Statistical patterns in triggered landslide event inventories and their use in a landslide-road impact model

Taylor, Faith Elizabeth

Awarding institution:
King's College London

Statistical Patterns in Triggered Landslide Event Inventories and their use in a Landslide-Road Impact Model

Faith Elizabeth Taylor

Thesis Submitted for the degree of Doctor of Philosophy (PhD)

Department of Geography, King's College London

30th April 2015

Abstract

This thesis examines landslide inventories, their statistical behaviour, and develops a generally-applicable landslide-road impact model exploiting this statistical behaviour. **Chapter 2** reviews landslide inventories, illustrating how production methods introduce uncertainty. **Chapter 3** applies a Boolean search to an archive of 568 UK regional newspapers to identify landslide information to add to the Great Britain *National Landslide Database*. Results for 2012 [2006] returned 73 [39] database additions. **Chapter 4** reviews generally applicable statistical models of triggered landslide event inventory area (A_L). **Chapter 5** investigates landslide shape using two large, substantially complete triggered-event inventories: 11,111 earthquake-triggered landslides (1994 Northridge, USA); 9,594 rainfall-triggered landslides (1998, Guatemala) and five additional 'lower confidence' inventories. Landslide polygons are abstracted to ellipses to calculate length-to-width ratios (L/W). Maximum-Likelihood Estimation bootstrapping techniques and Monte Carlo Kolmogorov-Smirnov testing show that an inverse-gamma pdf is a good general statistical model of landslide L/W when separated into categories of A_L. **Chapter 6** investigates spatial clustering in triggered-event inventories using pair-correlation to measure distances between all landslide centroid pairs and a moving-window technique to investigate linkages between landslide susceptibility and clustering. We find similarities in clustering across eight landslide inventories. **Chapter 7** exploits triggered landslide event statistics in a landslide-road impact model. Landslide areas and shapes are randomly selected from general statistical distributions and semi-randomly dropped over three regions (Northridge, USA; Collazone, Italy; Shu-Wa, Taiwan) conditioned by susceptibility. The resultant synthetic triggered-event inventories are overlaid with road maps and landslide road blocks identified. Model scenarios are run 100 times (Monte Carlo simulation) to create probabilistic forecasts, then confronted with observed triggered-events. By adjusting road corridor susceptibility, the model closely reflects observed road blockages numbers. This landslide-road impact model presents a low-data methodology to simulate simultaneous road network impacts.

Acknowledgements

Firstly, a bottomless thank you to my supervisor, Professor Bruce Malamud, who has spent many hours mentoring me to think about the underlying patterns of natural hazards, what it means to be a scientist and perhaps most importantly, how to communicate my research. Thanks also to my 2nd supervisor, Dr Andreas Baas for his continual support and contributing fresh perspectives at important stages of the research.

Throughout the PhD, I have been lucky enough to spend time with the fantastic team of scientists at the CNR-IRPI in Perugia, Italy. This includes Fausto Guzzetti, Michele Santangelo, Ivan Marchesini, Paola Reichenbach, Mauro Rossi, Alessandro Mondini and others. The CNR-IRPI has been an invaluable source of data, training about landslide inventory production, ideas and approaches to modelling. I am particularly grateful to Michele Santangelo, who visited us at KCL for 5 months in 2013 to work with me on modelling landslide road impact and Ivan Marchesini who was instrumental in the adoption of an open-source software approach.

Through Bruce Malamud and David Demerritt at KCL, I was given the opportunity to spend 6 months working with landslide researchers at the British Geological Survey and various bodies contributing to the UK Cabinet Office Natural Hazards Partnership. After mainly focusing on large, relatively catastrophic triggered landslide events abroad, this was an invaluable opportunity to learn about landslides closer to home and the policy approaches and challenges to dealing with natural hazards in the UK. Thank you to the NERC KTP for funding this internship.

This thesis would not have been possible without people sharing high quality landslide inventories and data. Thanks to the CNR-IRPI, EU FP7 LAMPRE Project, YiChin Chen at the Taiwan Typhoon and Flood Research Institute (NARL, Taiwan) and the United States Geological Survey for making inventories available. Thanks also to the community of open-source enthusiasts for tools and data like GRASS-GIS, Python, R and Open Street Map.

Getting paid to do research has been an absolute privilege. I am extremely grateful to the Natural Environment Research Council (NERC) for my PhD studentship, and the EU FP7 LAMPRE project for employing me as an associate scientist for the last few months of my PhD. I am also grateful to the EGU, ESA, King's College London, the AGU and Richard Hill Primary School Trust for providing grants which have allowed me to attend conferences and learn from a network of scientists, which has made the PhD experience richer.

My sincerest thanks go to the staff and students in Department of Geography at King's College London. I feel very lucky to work with such a vibrant community of passionate, driven and fun people – your energy is terrifying! Special thanks to Mark Griffiths, Benjamin Smith, Joel Gill, Anna Turbelin, Briony Turner, Kate Baker and the Intrepid Explorers team. Thanks also to former KCL postdoc Annette Witt (now at the Max Planck Institute for Dynamics and Self-Organization) for her comments and suggestions on the statistical approaches used in this thesis.

Many thanks to my viva examiners, Professor David Petley (University of East Anglia) and Professor Niels Hovius (GFZ Potsdam) for such valuable feedback on the thesis, and encouraging me to think beyond the immediate work I have performed here. It was an honour to have the time and input from two academics who have been so influential in shaping our current understanding of landslide hazard and risk.

Finally, the biggest thank you of all goes to my wonderful family (the terrific Taylors, brilliant Bulls and wonderful Weatherlys) and especially my partner, Andrew Weatherly. I will never forget the tremendous support, kindness, patience, proof-reading, emergency-telephone hotline attendance, much-needed distractions and amazing food you have given over the past four years (and before my PhD also). You are the best and I am forever grateful!

Table of Contents

CHAPTER SIX: SPATIAL CLUSTERING IN TRIGGERED LANDSLIDE EVENT INVENTORIES

PART III DEVELOPMENT OF A LANDSLIDE-ROAD IMPACT MODEL

CHAPTER SEVEN: THE LANDSLIDE-ROAD IMPACT MODEL

Index of Figures

CHAPTER 1

CHAPTER 2

CHAPTER 3

CHAPTER 4

CHAPTER 5

CHAPTER 6

Index of Tables

List of Acronyms and Abbreviations

Abbreviation	Meaning	Section Introduced
BGS	British Geological Survey	3.1
CH	Convex Hull	5.5
DEM	Digital Elevation Model	2.3
DInSAR	Differential Interferometric Synthetic Aperture Radar	2.4
E$f(A_{CH}, P_{CH})$	Quadratic ellipse (E) based on a function (f) of the area (A_{CH}) and perimeter (P_{CH}) of a convex hull (CH)	5.5
E$f((A_{CH}, P_{CH}) \rightarrow A_L)$	Quadratic ellipse (E) based on a function (f) of convex hull (CH) scaled (→) to landslide area (A_L)	5.5
E$f(A_L, P_L)$	Quadratic ellipse of landslide based on a function (f) of the area (A_L) and perimeter (P_L) of a landslide (L)	5.5
E_{SD}	Standard deviation (SD) ellipse	5.5
GIS	Geographic Information System	7.2
K-S	Kolmogorov-Smirnov Test	5.6
LiDAR	Light Detection and Ranging	2.4
LRIM	Landslide Road Impact Model	6.1
LS	Landslide	7.3
MLE	Maximum likelihood estimation	5.6
NHP	(UK Cabinet Office) Natural Hazards Partnership	3.1
NLD	National Landslide Database (of Great Britain)	3.1
pdf	Probability density function	4.2
SAR	Synthetic Aperture Radar	2.4
SD	Standard Deviation	4.2
SOC	Self-Organised Criticality	4.4
UAV	Unmanned Aerial Vehicle	2.4
UK	United Kingdom	3.1

List of Variables (Greek alphabetically then Roman alphabetically)

Variable	Meaning	Section Introduced
α	Scale parameter for double Pareto pdf	4.3
α	Alpha measure of network connectivity	7.5
β	Parameter that controls the power-law gradient of the left-hand tail for double Pareto pdf	4.3
Γ	Gamma function ($\Gamma(x) = (x-1)!$) Where ! = factorial	4.3
δ	Bin width (right most minus left most boundaries of the bin)	4.2
ϵ	Number of decimal places precision for significance level of Monte-Carlo Kolmogorov-Smirnov goodness-of-fit test	5.6
$\varkappa_{AR}$	Normalised area of road (AR) network that has been buffered by 100 m either side of the road line, per km^2	7.5
$\varkappa_{BL}$	Normalised number of road blockages (BL) per km^2	6.2
$\varkappa_{LS}$	Normalised number of landslide (LS) centroids per km^2	6.2
μ	Location parameter (mean) for normal and log-normal pdf	4.2
ρ	Scale parameter for inverse Gamma pdf	4.3
σ	Standard deviation parameter for normal and log-normal pdf	4.2
ϕ	Soil friction angle	7.4
a	Shape parameter for inverse Gamma pdf	4.3
A_C	Area of a grid cell (C)	6.2
A_{CH}	Area of best-fit convex hull (CH) to a polygon	5.5
A_{Circ}	Area of a circular (Circ) study region	6.2
A_E	Area of ellipse	5.5
A_I	Area of intersection between landslide polygon and elliptical approximation	5.5
A_L	Individual landslide area	2.3
A_S	Area of study region extent	6.2
c	Cut-off' parameter controlling the minimum value for double Pareto pdf	4.3
c	Soil cohesion	7.4
D	Date precision of landslide record (number of days either side of date in which landslide could have occurred)	3.4
D	Distance from landslide centroid to nearest road	7.3
D_{OBS}	Kolmogorov-Smirnov test: distance between observed cumulative frequency (OBS) and MLE fit cumulative frequency curves	5.6

Variable	Meaning	Section Introduced
D_R	Damage to road (R)	7.2
D_{SIM}	Kolmogorov-Smirnov test: distance between simulated data cumulative frequency (SIM) and MLE fit cumulative frequency curves	5.6
e	Ellipticity goodness-of-fit measure	5.5
E	Number of edges in a spatial network	7.5
$f(x)$	Frequency density of x	4.2
$F(x)$	Cumulative distribution function of x	5.6
$F_n(x)$	Empirical cumulative distribution function of x	5.6
g	Soil density	7.4
k	Power-law constant for power-law distributions	4.3
k	Hydraulic conductivity	7.4
L	Time lag between landslide occurrence and reporting (variable units)	3.4
L	Length	5.3
L/W	Length-to-width ratio	5.2
L_D	Average landslide density across a region	6.4
L_{RD}	Length of road (RD)	7.3
m	Cut-off' parameter controlling the maximum value for the Pareto pdf	4.3
M_w	Earthquake moment magnitude	2.4
n	Number or count (different variable types)	4.2
N	Number of nodes in a network	7.5
(N, E)	Northing and Easting coordinates	6.1
N_{BL}	Number of road blockages (BL)	6.5
N_{Cells}	Total number of 1 km × 1 km grid cells required to cover a study region	6.2
n_{iter}	Number of iterations (iter) for Monte-Carlo Kolmogorov-Smirnov goodness-of-fit test	5.6
N_{LS}	Number of landslides (LS)	2..4
p	Kolmogorov-Smirnov goodness-of-fit test significance level	5.6
p	Number of isolated subgraphs in a network	7.5
$p(x)$	Probability Density of x	4.2
P_{CH}	Perimeter of convex hull (CH)	5.5
P_E	Perimeter of ellipse (E)	5.5
P_L	Perimeter of landslide (L)	5.5

Variable	Meaning	Section Introduced
r	Linear distance between pairs of centroids	6.2
S	Spatial precision of a landslide record (estimated radius from true landslide location in metres)	3.4
s	Location parameter for inverse Gamma pdf	4.3
S_{LS}	Landslide (LS) susceptibility	6.1
Sup	Supremum (maximum) of a set of values	5.6
t	Parameter that controls the location of the maximum probability for double Pareto pdf	4.3
W	Width	5.3
x	Random real uniform value	6.1
z	Vertical thickness of soil	7.4

List of Equations

Equation Number	Equation	Section Introduced
4.1	Calculation of frequency density	4.2
4.2	Calculation of probability density	4.2
4.3	Integral of probability density function	4.2
4.4	Gaussian probability density function	4.2
4.5	Log-normal probability density function	4.2
4.6	Inverse power-law probability density function	4.3
4.7	Inverse Gamma probability density function	4.3
4.8	Double Pareto probability density function	4.3
5.1	Area of an ellipse	5.5
5.2	Approximation of perimeter of an ellipse	5.5
5.3	Calculation of ellipse length-to-width ratio (L/W) from ellipse area and perimeter	5.5
5.4	Ellipticity goodness-of-fit measure	5.5
5.5	Inverse-Gamma probability density function	5.6
5.6	Kolmogorov-Smirnov goodness-of-fit test	5.6
5.7	Number of iterations for Monte-Carlo Kolmogorov-Smirnov goodness-of-fit test	5.6
7.1	Inverse Gamma probability density function	7.3
7.2	Alpha measure of road network connectivity	7.5

For my brother Frank Taylor,

What is mine is yours.

Chapter 1: Introduction

"And I sure love the sound of the rain pouring down on my carport roof made out of tin.
If there's a flood then there's gonna be mudslides, we all have to pay for our sin.
And I suppose they'll close canyon roads, and the freeways will all start to clog.
And the waters will rise and you won't be surprised when your whole house smells like a wet dog"
From *Grey in L.A.* by Loudon Wainwright III (1946–)

1.1 Introduction

Triggers such as earthquakes or heavy rainfall can result in single to thousands of landslides occurring across a region within a short space of time (Crozier, 2005). This is illustrated in in **Figure 1.1**, where an oblique photograph shows many hundreds of the more than 10,000 landslides that were triggered across the 49 km^2 Hawke's Bay region of New Zealand by heavy rain as part of Cyclone Bola in March 1988 (Glade, 2001). Each of these landslides that are part of this triggered landslide event (where an event consists of the 'trigger' plus the single to thousands of landslides resulting from the trigger), has the potential, depending on the landslide's location, to block parts of the road network, impacting how people move about a region and slowing recovery efforts.

Figure 1.1 *Landslides triggered by Cyclone Bola in March 1988, New Zealand. Each light coloured area is a landslide. Photography from Trustrum (1988).*

Many regions particularly susceptible to large triggered landslide events are in low- to middle-income countries, where vulnerability is high and technical capacity to forecast and

deal with events may be low (Guinau *et al.*, 2005; Yang *et al.*, 2015). Yet as we will show in the literature reviewed throughout this thesis, much of our understanding and ability to forecast these triggered landslide events depends upon high quality, detailed observations of local conditions and past events, which are often difficult, time consuming and expensive to collect. Thus, the over-arching question that guides the research performed in this thesis is "*how can we work around the issue of low data confidence for understanding landslide risk and impact?*". We aim to contribute to the knowledge of this topic through three different methodologies:

1. *Developing a robust and repeatable methodology to obtain records of past landslides and their impact from a digital archive of regional newspapers* to supplement a landslide database with more records and richer information.
2. *Measuring the statistical behaviour of high quality, substantially complete triggered landslide event inventories* and testing whether this might be 'generally applicable' and thus used as a 'template' for triggered landslide event behaviour in locations where high quality data is not available.
3. *Exploiting generally applicable statistical behaviour of triggered landslide event inventories in a semi-stochastic, simple model* that creates synthetic triggered landslide event inventories which we use to forecast road network impact caused by landslides.

These three ways of addressing the issue of low data availability form the three major parts of the thesis, each part consisting of background literature review chapters and primary research chapters. In the following **Section 1.2** we discuss the rationale for this 'low data' approach by illustrating how availability of high quality and relevant data is often a limiting factor in calculating landslide risk.

1.2 Rationale

1.2.1 What is Landslide Risk and Why is it Difficult to Quantify?

In its simplest form, landslide risk represents the probability of loss of a specific element (e.g., infrastructure, loss of life) within a given time period, which is calculated from the product of landslide hazard (the spatial and temporal probability of a landslide occurring) and the consequence of a landslide impacting an element (e.g., the amount and cost of damage) (Dai *et al.*, 2002). However, van Westen *et al.* (2006) argue that this risk calculation is deceptive in its simplicity, and the calculation of total risk requires the analysis of many parameters, which are each subject to uncertainties. **Figure 1.2** shows van Westen *et al.*

(2006)'s schematic representation of the data and methodologies required to calculate total landslide risk, illustrating at least 14 input variables and 6 methodologies.

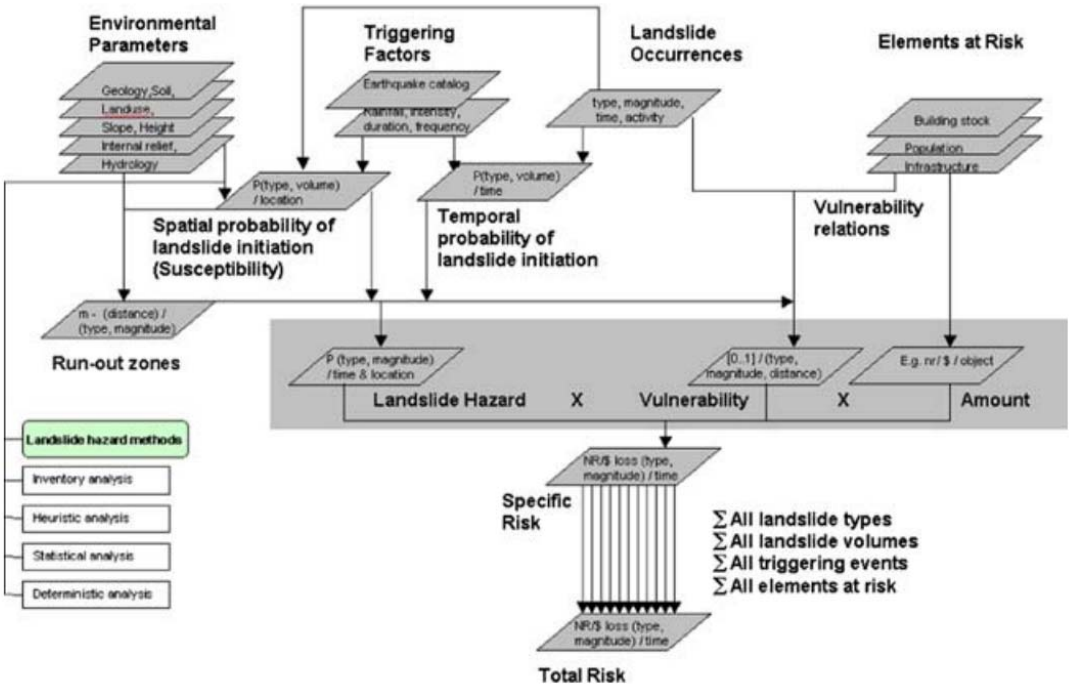


Figure 1.2 *Variables and methodologies required to calculate total landslide risk for a given area. Figure from van Westen et al. (2006).*

Guzzetti *et al.* (1999) note that in reality, many assessments of landslide hazard (excluding the confounding factor of calculating the consequences to human lives and property) are partial or uncertain due to issues of:

1. *Poor or partial records* of past landslide events and their causes.
2. *Access to quality datasets of local conditions* (e.g., geology, climate).
3. *Selection of appropriate techniques* and tools to calculate landslide hazard.

We will now discuss the above three points from Guzzetti *et al.* (1999) and highlight how this thesis aims to contribute to addressing these issues.

1.2.1.1 Issues of Poor or Partial Records of Past Landslide Events

Related to Guzzetti *et al.* (1999)'s point 1 above (*poor or partial records*), it is well established that records of past landslide events are the most important factor in understanding landslide risk (van Westen *et al.*, 2008; Guzzetti *et al.,* 2012; Van Den Eeckhaut and Hervás, 2012). Indeed, more broadly the collection of standardised relevant data on hazards is a major priority of the recent Sendai Framework for Disaster Risk Reduction 2015—2030 (UNISDR, 2015). Yet, as we will discuss in more detail in **Chapter 2**, databases of past

landslide occurrence and impact are often incomplete (Wills and McCrink, 2002; Van Den Eeckhaut and Hervás, 2012; Guzzetti *et al.*, 2012), which results in uncertain or biased statistical analysis and assessment of hazard and risk. This incompleteness stems from issues such as the following:

- *Small landslides being erased from the landscape* through process of erosion and anthropic activity (Malamud *et al.*, 2004a; Bell *et al.,* 2012).
- *Lack of standardised and shared data collection* between various bodies involved in landslide management (Wills *et al.*, 2014).
- *Landslide impacts being recorded as an impact from the triggering event* (e.g., earthquake or hurricane) (Kirschbaum *et al.,* 2010).

Part I of this thesis (**Chapters 2** and **3**) aims to (i) review the literature regarding how landslide inventories and databases are created and their sources of uncertainty (**Chapter 2**) and (ii) present a method to deal with some of these shortfalls in records of past landslide events by establishing a robust and repeatable method to search a digital archive of newspapers to find more records of landslides and richer information about landslides.

1.2.1.2 Issues of Access to Quality Datasets of Local Conditions

Related to Guzzetti *et al.* (1999)'s point 2 above (*access to quality datasets of local conditions*), the collection of 'baseline' data of local conditions is frequently a low priority for disaster managers, compared to 'reactive' data collection during the response to hazard event (Amin *et al.*, 2008). Yet, as illustrated by van Westen *et al.* (2006) in **Figure 1.2**, many baseline environmental datasets are required to calculate landslide hazard, typically including: geology, soil, land use, slope and hydrology. In **Part II** of this thesis (**Chapters 4–6**), we aim to circumvent the issue of detailed data collection by establishing 'generally applicable' statistical behaviour of triggered landslide event inventories from locations where high quality data are available, which can then be applied to locations where it is not feasible to collect environmental datasets.

This generally observed triggered landslide event behaviour is (i) already established landslide frequency-area statistics (e.g., Stark and Hovius, 2001; Malamud *et al.,* 2004a) (reviewed in **Chapter 4**), (ii) primary research to quantify the shape of landslides in triggered landslide event inventories (**Chapter 5**) and (iii) primary research to quantify the spatial clustering of landslides in triggered landslide event inventories (**Chapter 6**). Until recently, it appears that the majority of the literature investigating the statistical behaviour

of triggered landslide events has been more concerned with the scientific aspects of methods of analysis and physical explanations for observed behaviour (e.g., Pelletier *et al.,* 1997; Stark and Hovius, 2001; Malamud *et al.,* 2004a; Katz and Aharonov, 2006; Stark and Guzzetti, 2009; Katz *et al.,* 2014). Consequently, the methodology developed here not only contributes to the body of literature looking at triggered landslide event inventory behaviour from a statistical and scientific perspective, but also demonstrates new ways to exploit this statistical behaviour in an applied sense for civil protection purposes and risk assessment.

1.2.1.3 Issues of Appropriate Techniques and Tools to Calculate Landslide Hazard

Related to Guzzetti *et al.* (1999)'s point 3 above (*selection of appropriate techniques and tools to calculate landslide hazard*), van Asch *et al.* (2007) argue that there have been great advances in our understanding of and ability to deterministically calculate landslide hazard when the above appropriate datasets and tools are available, but the real challenge in advancing our understanding of landslide hazard is developing 'adequate' models that do a 'good enough' job with the data available.

The development of 'adequate' models is particularly relevant when considering triggered landslide event impact upon road networks as we do in **Part III** (**Chapter 7**) of this thesis. This is because many regions where landslide hazard is high and road network vulnerability is high are in low- to middle-income regions where resources may not be available for the systematic collection of baseline and past landslide incidence data. This is illustrated in **Figure 1.3** where we show three maps:

- *Human development index* (**Figure 1.3A**) as a proxy for capacity to collect and maintain relevant datasets for analysis of landslide hazard.
- *Over-ground travel time to nearest city* (**Figure 1.3B**) as a proxy for road network vulnerability and risk of isolation when landslides cause road blockages.
- *Landslide susceptibility* (**Figure 1.3C**) as a proxy for regions where there is a high probability of large triggered landslide events occurring.

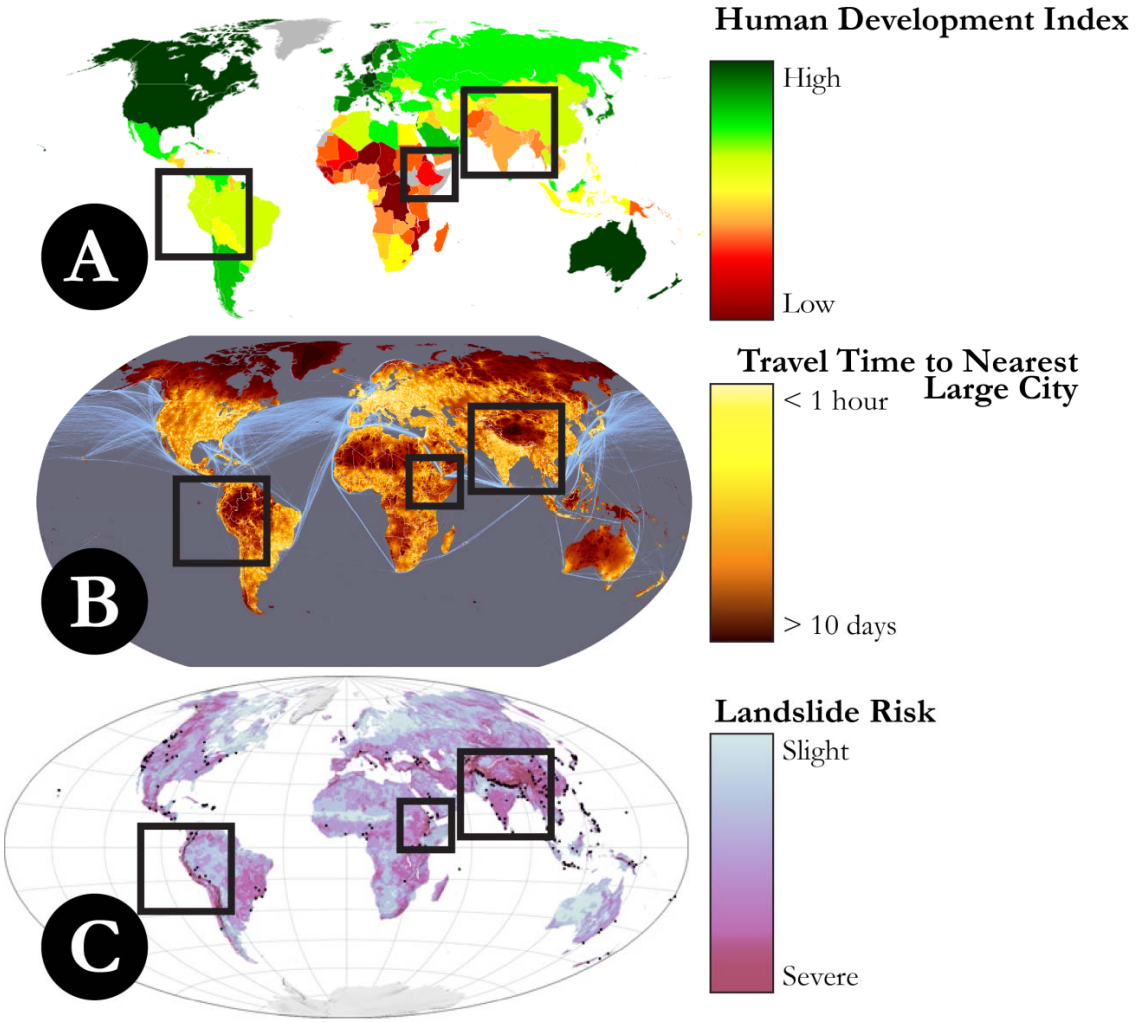


Figure 1.3 *Illustration of the concept that many regions where landslide risk is high are in the Global South in terms of economic status (using the Human Development Index as a rough indicator). These regions are also often quite remote in terms of overland travel time to nearest large city.* ***(A)*** *Map of human development index which is a measure of quality of life, in terms of life expectancy, education and economic status (United Nations, 2014),* ***(B)*** *Map of travel time to the nearest city by land (Uchida and Nelson, 2009),* ***(C)*** *Map of global landslide risk, which indicates the relative likelihood of observing a landslide at a given spatial location (Hong et al., 2007). Three regions have been highlighted (black boxes) where these three factors intersect to create high vulnerability (both human and road network) to triggered landslide events. These are: the Andes mountain range in South West South America (Colombia, Ecuador, Peru, Bolivia); the Ethiopian Highlands in the Horn of Africa and the Himalayan mountain range across Nepal, Pakistan and Bhutan.*

Figure 1.3 illustrates that there are many regions across the world where the factors of high landslide hazard and high road network vulnerability intersect with low technical capacity available to collect data, forecast and respond to the risk of landslide-road impact.

Our current understanding of and ability to forecast the impact of triggered landslide events upon regional networks is largely driven by the following techniques:

(i) *Forensic cost studies and economic models* where the cost of previous events is extrapolated or scaled to forecast the regional impact of future events (e.g., Wills *et al.*, 2014).

(ii) *Overlaying observed landslide inventory maps on road network maps* (e.g., Guzzetti *et al.,* 2003).

(iii) *Overlaying landslide susceptibility maps with regional road network maps to identify high risk stretches of road* (e.g., Remondo *et al.*, 2008).

The above three techniques often require high volumes of data and local parameterisation, and do not typically examine the potential for multiple simultaneous road blockages. Note that here, we use the term 'simultaneous' to denote a set of road blockages across a region caused by landslides occurring in the minutes to weeks after a triggering event. We acknowledge that it is possible that some landslide road blockages may be remediated whist other road blockages are occurring, but we use the term 'simultaneous' loosely to describe the regional impact of landslides blocking the road in the weeks to months following a large triggering event.

The impact of simultaneous road blockages may be greater than the sum of its parts if settlements become isolated by road. For example, if other 'lifeline' infrastructures are damaged (e.g., communication, fresh water, power), roads are particularly crucial for bringing in supplies (Cova and Conger, 2004). But if key access routes are blocked, immediate relief necessities such as fresh water, medical supplies, generators, search teams and personnel cannot easily reach the affected and indeed the fatalities of the hazard may be prolonged and worsened (Haghani and Oh, 1996, Kuwata and Takada, 2004). Schulz *et al.* (2012) also note the importance of multiple hazard interactions, discussing the potential for evacuation routes to become blocked by earthquake-triggered landslides, which could potentially expose humans to additional earthquake-triggered tsunami risk.

Recent triggered landslide events in the media include the Sikkim Earthquake (September 2011), where due to the heavy rain following the event, coupled with many road blockages caused by landslides, it took four days to reach some of the worst affected regions even by helicopter (OCHA, 2011) and some villages remained isolated for a period after this. In the Wenchuan Earthquake (May 2008), 6,074 deaths were directly caused by landslides of which more than 1,000 were estimated to have been where landslides occurred on the road (Yin et al., 2009). In this case, the number of deaths indirectly caused by landslides blocking relief efforts to a region for days-months might have been considerably higher. Gordon *et al.* (1998) estimated that the total economic loss resulting from transportation disruption following the 1994 Northridge earthquake in California, USA was in excess of $1.5 thousand million. Blockage of key access routes may also slow a region's long term

recovery if, for example, heavy machinery cannot be brought in for rebuilding, resulting in knock-on economic, social and psychological implications for those affected by the hazard (Oven, 2009; Sökefeld, 2012).

Here we hope to develop a tool that does the following:

(i) *requires few regionally specific data inputs*

(ii) *simulates multiple simultaneous road blockages* caused by landslides during a triggered landslide event

(iii) *is generally applicable* to many regions of the world susceptible to triggered landslide events

(iv) *is created using open source software* which anyone can use and modify

With this tool we hope to address some of the shortcomings of current techniques used to forecast landslide-road network impact and improve understanding of landslide risk in regions where data scarcity is problematic.

1.3 Research Questions

This thesis is focused on methodologies to (i) create richer databases of landslides, (ii) measure statistical behaviour of triggered landslide event inventories and test whether this is 'generally applicable' and (iii) model synthetic triggered landslide event inventory maps and their impact upon road networks by exploiting 'generally applicable' statistical behaviour of triggered landslide event inventories. In focussing on these methodologies, the following four research questions are addressed:

1. Is it possible to obtain richer information about the occurrence and impacts of landslides in Great Britain from databases of regional newspapers?
2. What shape can landslides be modelled as, and how does this shape vary? Question sub-parts include:
    a. Can we abstract landslide shapes to ellipses in a robust, repeatable way?
    b. Does landslide ellipse length-to-width ratio vary with landslide size?
    c. Is the statistical behaviour of landslide shape 'generally applicable' across different landslide inventories from different locations?
3. How can the spatial clustering of landslides in triggered event inventories be measured and is this 'generally applicable' behaviour across different landslide inventories from different locations?

4. Is it possible to create a 'generally applicable' model of triggered landslide event impact upon regional road networks using statistical behaviour of landslide size and shape? Question sub-parts include:
    a. How can we generate synthetic triggered landslide inventory maps from statistical behaviour of landslides?
    b. Do local approaches to road building and landslide protection affect landslide-road interactions?
    c. Can this model be applied to a variety of locations?

The chapters in which these research questions are answered are outlined in the following.

1.4 Thesis Outline

This thesis is organised into three major parts, each containing a background literature and methods review (either as a separate chapter or as part of the chapter) and primary research:

(i) **Part I:** Landslide inventories and completeness (**Chapters 2 and 3**)
(ii) **Part II:** Statistical models of landslide size, shape and clustering (**Chapters 4–6**)
(iii) **Part III:** Development of a landslide-road impact model (**Chapter 7**)

A summary of each chapter is presented below:

PART I: LANDSLIDE INVENTORIES AND COMPLETENESS

Chapter 2: Landslide Inventories and Databases. This background literature and concepts chapter reviews (i) what different types of landslide inventory exist, (ii) methods of inventory production and (iii) their associated uncertainties that may result in under-sampling of certain types of landslide. It then discusses why inventory completeness is important in the statistical analyses of landslide size, shape and location (performed in **Section 2** of this thesis) to ensure a representative sample of all landslides in terms of size, shape and clustering. A criteria of inventory 'completeness' is defined, which is used to select inventories analysed in **Sections 2 and 3** of this thesis.

Chapter 3: Adding richness to the National Landslide Database of Great Britain. This primary research chapter develops a robust Boolean search and applies it to a digital archive of 568 UK regional newspapers to identify landslide articles and impact information. Results for 2012 [2006] returned 1,668 [711]

articles, with 73 [39] additions to the *National Landslide Database* of Great Britain (maintained by the *British Geological Survey*) and shows how archival methodologies can enhance landslide database richness in terms of number of landslides recorded and richness of information for each landslide.

PART II: STATISTICAL MODELS OF LANDSLIDE SHAPE, SIZE AND CLUSTERING

Chapter 4: A Review of Statistical Models for Landslide Area. This background literature and concepts chapter reviews and provides a methodological background for statistical models of triggered landslide event landslide areas. The concepts of frequency and probability density are briefly explained, followed by a review of models used to describe the probability distribution of landslide areas in different triggered landslide event inventories. The concept of 'generally applicable' statistical distributions to model landslide area across many different regions is introduced, followed by a review of different mechanisms put forward in the literature to explain this generally observed behaviour.

Chapter 5: What Shape is a Landslide? Statistical Distributions of Landslide Length to Width Ratio. This primary research chapter presents a methodology to quantify the shape of landslides by measuring variability in landslide length (L) to width (W) ratio. We use two large, substantially complete triggered-event inventories available publically: 11,111 earthquake-triggered landslides (1994 Northridge, USA) and 9,594 rainfall triggered landslides (1998 Hurricane Mitch, Guatemala). We experiment with four methods to abstract landslide polygons to ellipses and find the best method is obtained by fitting a convex hull (CH) to each landslide polygon, approximating an ellipse with the equivalent area (A_{CH}) and perimeter (P_{CH}) of the convex hull and then scaling this ellipse to match the area of the original landslide (A_L). The goodness-of-fit (e) of elliptical approximations was calculated using a measure of the area of intersection (A_I) between the original landslide polygon area (A_L) and the elliptical approximation. We find that > 80% of landslides can be reasonably well approximated by an ellipse in both inventories. Length-to-width ratios (L/W) of ellipses are calculated for all landslides with elliptical goodness-of-fit $e \geq 0.5$. The statistical distribution of length-to-width (L/W) ratios are then examined, both for all landslide areas (A_L) and subcategories of landslide areas. Using Maximum-Likelihood Estimation bootstrapping techniques and Monte Carlo Kolmogorov-Smirnov testing, we find that an inverse-

gamma probability density function is a good statistical model for the probability of landslide L/W, with parameters of the inverse-gamma pdf depending on the landslide area category considered. The methodology is then applied to five additional 'lower confidence' triggered landslide event inventories with 149–851 landslides in each inventory, to confront whether landslide shape exhibits 'generally applicable behaviour'. There is some similarity in L/W behaviour in landslide inventories where the trigger involves water, where L/W typically increases with landslide area (A_L). For the one earthquake-triggered landslide inventory (Northridge), L/W tends to decrease with A_L. This suggests a potential 'semi-generally applicable' behaviour of landslide shape, depending on the triggering mechanism, although there is uncertainty due to the small sample sizes in the lower confidence inventories.

Chapter 6: Spatial clustering in triggered landslide event inventories. This is a short primary research chapter presenting results of clustering when developing the landslide-road impact model (presented later in **Chapter 7**). Using eight triggered landslide event inventories varying in number of landslides, location, triggering event and confidence in quality of the inventory, we use pair-correlation to measure linear distances (r) between all landslide centroid pairs, and analyse the probability density distribution ($p(r)$) of the distances between all pairs. For all eight inventories examined, $p(r)$ differs from the probability density distribution of a randomly generated set of points, particularly at scales $r < 1.2$ km – 40.9 km (depending on the size of the study region), showing that triggered landslide event inventories exhibit spatial clustering at small scales. There are similarities in $p(r)$ across all eight inventories when normalised by the area of the study region, although inventories dominated by 'low-mobility' landslides (where the majority of landslides are low-run out) tend to exhibit clustering at a finer spatial scale than inventories dominated by 'high mobility' landslides (where the majority of landslides are debris flows and rockfalls). In the second half of **Chapter 6**, we develop a method (the Kappa-measure, κ) to measure the number of landslide centroids per square kilometre of a moving-window grid, normalised by the proportion of the grid cell covered by study area. This is applied to two substantially complete triggered landslide event inventories (Collazzone, Central Italy where 413 landslides were triggered by snowmelt in 1997, and Oat Mountain, Northridge, USA where 1,356 landslides were triggered by the 1994 Northridge Earthquake) where landslide susceptibility

maps were also available. Using the kappa-measure, we show that the link between landslide susceptibility (S_{LS}) and number of landslides per square kilometre is not straightforward. That is, some low landslide susceptibility zones ($S_{LS} < \approx 0.2$) fail multiple times (up to 10 landslides per square kilometre), and some high landslide susceptibility zones ($S_{LS} > \approx 0.8$) will not fail at all in a particular triggered landslide event (i.e., there are 0 landslides per square kilometre). We test whether the inclusion of spatial clustering of landslides in a landslide-road impact model improves model skill in forecasting landslide-road impact and find this does not significantly improve results.

PART III: DEVELOPMENT OF A LANDSLIDE-ROAD IMPACT MODEL (LRIM)

Chapter 7: The Landslide-Road Impact Model (LRIM). This primary research chapter outlines the development and application of a landslide-road impact model (LRIM) to simulate triggered landslide events and their impacts upon road networks. This model exploits the statistical behaviour of triggered landslide event inventories established in **Section 2** of this thesis, and after application in study regions is then confronted with 'real' landslide inventories. The LRIM creates 'synthetic' triggered landslide event inventories by randomly sampling landslide areas and shapes from already established statistical distributions (described in **Chapter 4** and established in **Chapter 5**, respectively). In the methodology, these landslides are then semi-randomly dropped across a study region, conditioned by a landslide susceptibility map. The resulting synthetic triggered landslide event inventory is overlaid with a road network map and the number, size, location and network impact of road blockages calculated. This process is repeated hundreds of times in a Monte Carlo type simulation. Because the statistical distributions and approaches used in the model are thought to be generally applicable for low-mobility triggered landslides in many regions of medium- to high-topography throughout the world, relatively little detailed local data is required to run the model. Coupled with an open-source modelling approach, this model may be applied to many regions where triggered landslide events occur. We apply LRIM, and confront with observed data for three study regions: (i) Collazzone (Central Italy) where rapid snowmelt triggered 413 landslides in January 1997; (ii) Oat Mountain (Northridge, USA), where the Northridge Earthquake triggered 1,356 landslides in January 1994; (iii) Su-Hua (northeast Taiwan) where Typhoon Megi

triggered 149 landslides in 2010. We find that when the landslide susceptibility map is adjusted along road corridors to take into account interactions between landslides and roads, the model matches reasonably well the observed results. In Collazzone (length of road = 153 km, landslide density = 5.2 landslides km^{-2}), the median over 100 model runs was 5 (±2.5 standard deviation (S.D.)) road blocks, compared to the observed number of 5 road blocks. In Oat Mountain (length of road = 780 km, landslide density = 8.7 landslides km^{-2}), the median over 100 model runs was 108 (±17.2 S.D.) road blocks compared to the observed number of 48 road blocks. In Su-Hua (length of road = 265 km, landslide density = 0.35 landslides km^{-2}), the median over 100 model runs was 8 (±3.8 S.D.) road blocks compared to the observed number of 18 road blocks. This landslide-road impact model presents an open-source low-data methodology for the stochastic simulation of potential simultaneous road network impacts.

Chapter 8: Summary and Conclusions This chapter summarises and draws together findings from the background and research chapters. We address how we have answered our original research questions (**Section 1.3**) and contributed to the scientific knowledge on dealing with the issue of low data availability for understanding landslide incidence and impact and establishing generally applicable behaviour of triggered landslide event inventories.

Appendix A: R and Python codes to measure and statistically analyse the shape of landslides. The open-source codes developed to analyse the shape of landslides in **Chapter 5** are presented.

Appendix B: Python code for the Landslide-Road Impact Model. The open-source python code for the landslide-road impact model (LRIM) in **Chapter 7** is presented.

1.5 Conclusions

This thesis develops methodologies and tools that are (i) open-source and freely available for others to use and apply in their own contexts and (ii) thought to be reasonably 'generally applicable' to many regions susceptible to low-mobility triggered landslide events. Through the development of these methodologies and tools, we hope to contribute to the understanding and ability to forecast triggered landslide events in regions where there is low data availability.

PART I: LANDSLIDE INVENTORIES AND COMPLETENESS

Chapter 2: Landslide Inventories and Databases

Chapter 2 Summary

This **Chapter 2** is primarily (i) a literature review of landslide inventory and database taxonomy; (ii) a summary of methods of landslide inventory/database production; and (iii) a discussion of uncertainties encountered when compiling landslide inventories, which has implications for the statistical analysis and modelling performed throughout this thesis. We find that the terms landslide 'inventory' and 'database' are not clearly defined in the literature and sometimes used interchangeably. We define a landslide database as a collection of records of landslides and an inventory as aiming towards a 'complete' record of a particular 'set' of landslides, although completeness is rarely possible. We review how landslide inventories can be classified by the method of production, the age of landslides included and the scale of mapping. We briefly explain five principle methods used to catalogue landslides in inventories and databases, including: field techniques, aerial photo interpretation, novel ground and airborne remote sensing techniques, satellite remote sensing and archive methods. For each method, we discuss potential uncertainties, particularly in terms of inventory 'completeness' and explain how there is often a bias towards recording large size landslides, as these are more easily discernible and less rapidly erased from the landscape than small size landslides. We discuss critically the concept of landslide inventory and database completeness and set out criteria for landslide inventories suitable for statistical analysis. We explain that well-produced triggered landslide event inventories tend to be more representative of the range of landslide sizes, shapes and spatial locations as all landslide scars are 'fresh' and thus have not been erased from the landscape. These criteria are used in **Chapters 4–7** to select landslide inventories which we statistically analyse and model.

*List of abbreviations used in **Chapter 2***

Abbreviation	Meaning	Section Introduced
DEM	Digital Elevation Model	2.3
DInSAR	Differential Interferometric Synthetic Aperture Radar	2.4
LiDAR	Light Detection and Ranging	2.4
SAR	Synthetic Aperture Radar	2.4
UAV	Unmanned Aerial Vehicle	2.4

*List of variables used in **Chapter 2***

Variable	Meaning	Section Introduced
A_L	Individual landslide area	2.3
M_w	Earthquake moment magnitude	2.4
N_{LS}	Number of landslides	2..4

2.1 Introduction

Landslide inventories and databases are investigated throughout this thesis for:

- *Developing methods to add richness to landslide a database* by searching archives of local newspapers (**Chapter 3**).
- *Reviewing and analysing statistical patterns* exhibited by triggered landslide events in terms of landslide size (**Chapter 4**), shape (**Chapter 5**) and spatial clustering (**Chapter 6**).
- *Exploiting these statistical patterns for simulating triggered landslide events* in the Landslide Road Impact Model (LRIM) (**Chapter 7**).
- *Confronting LRIM output* with real triggered landslide event inventories (**Chapter 7**).
- *Production of landslide susceptibility maps* (performed by others but used in **Chapter 6** and **Chapter 7**).

For each of the above tasks, it is important for us to understand the nature and limitations of the landslide data we are working with, particularly in terms of database or inventory 'completeness'. In this **Chapter 2**, we critically review the literature concerned with landslide inventories and databases, which forms an evidence base for the analysis performed in **Chapters 3**, **5, 6** and **7** and a background for the review of landslide frequency-size statistics in **Chapter 4**. The following sections of this **Chapter 2**:

2.2 Explain what are landslide databases and inventories.

2.3 Describe how landslide inventories are classified into types and give examples of how the type of inventory will affect what information can be extracted from that dataset.

2.4 Discuss different methods of inventory production and their associated uncertainties.

2.5 Discuss the implications of findings in **Sections 2.2–2.4** on the analyses performed later in this thesis.

2.6 Summarise and conclude the chapter.

2.2 What are Landslide Databases and Inventories?

As is the case for many natural hazards, records of past occurrence, characteristics and impacts underpin our understanding of the spatiotemporal landslide hazard and risk (Schuster and Kockelman, 1996; van Westen *et al.,* 2008). For many natural hazards,

records of past occurrence are compiled from direct instrumental measurements. For example: water depth for floods (e.g., Bayliss and Jones, 1993); magnitude of ground shaking for earthquakes (e.g., USGS, 2015); wind speed for hurricanes (e.g., Landsea *et al.,* 2004). However, a landslide deposit (and associated erosional surfaces) observed on the ground are the *outcome* of a set of interacting processes (Guzzetti *et al.,* 1999) that are rarely feasible to measure systematically and instrumentally. Consequently, to create a record of landslides, one must actively search for them across a landscape, through methods such as remote sensing and photogrammetry, field investigations, archival research or a combination thereof (Guzzetti *et al.,* 2012). Thus a landslide inventory or database represents a spatial record of landslides that have either left discernible features (Soeters and van Westen, 1996) or that humans are/were aware of (Guzzetti *et al.,* 1994). Landslide inventories and databases are produced for a variety of purposes, which Guzzetti *et al.* (2012) categorise as:

(i) *The documentation of the extent* of landslides across a region.

(ii) *The creation of landslide susceptibility* and hazard assessments.

(iii) *The investigation of the distribution, types, and patterns of landslides* in relation to morphological and geological characteristics.

(iv) *The understanding of the evolution of landscapes* dominated by mass wasting processes.

(v) *The informing of the rescue and recovery* processes.

The literature is not always clear on the distinction between a landslide 'database' and a landslide 'inventory'. Indeed, the terms are sometimes used interchangeably (van Westen *et al.,* 2008; Hervás, 2013; Damm and Klose, 2015), or landslide databases have been described as 'digital' inventories (van Den Eeckhaut and Hervás, 2012). However, we find it useful to make a semantic distinction based on the dictionary definitions from the Oxford Dictionary (2015a,b, no pagination):

> "**Database:** A structured set of data held in a computer."
>
> "**Inventory:** A complete list of items."

Based on these two definitions, we term a landslide database as a collection of records of individual landslides (although not necessarily digital), whereas an inventory *aims towards* being a *complete* record of a particular 'set' of landslides (where that 'set' could be: landslides over all time, landslides from an individual trigger, landslides resulting from a particular activity, etc.). Although, in reality, it is unlikely that any inventory is truly complete for reasons we discuss later in **Section 2.4**. Based on the research performed throughout this

thesis using landslide inventories and databases, and existing literature (Dikau *et al.*, 1996; Petley *et al.,* 2005; van Den Eeckhaut and Hervás, 2012; Guzzetti *et al.,* 2012), we believe the types of information in a landslide database or inventory are fairly similar, and they will typically include the following information fields:

- *Landslide location* (e.g., coordinates, place name, region).
- *Landslide extent* (e.g., area, volume, perimeter).
- *Temporal information* (e.g., date of occurrence, date of last activity).
- *Geological information* (e.g., landslide type, landslide trigger, what part of the landslide is mapped, such as crown, scarp, etc.).
- *Impacts* (e.g., fatalities, injuries, economic impact).
- *Management* (e.g., information about how the landslide was surveyed, any remedial measures taken).
- *Metadata* (e.g., source of inventory information, who produced the inventory, estimate of uncertainty).

Although all of the above fields may be present in a landslide database or inventory, it is not always possible to record all information about a landslide due to methods used; intended use of the inventory/database; data limitations; age of the landslide and uncertainties in quantifying impact (Petley *et al.,* 2005), which we discuss in **Section 2.4**. By our distinction, a landslide database could contain various different inventories, or be list of 'secular' landslides, or a combination of both. In the following **Section 2.3**, we focus on what different types of landslide inventory exist. We emphasise here that many landslide databases (which may be referred to as an inventory by the author) will be comprised of various different types of inventory.

2.3 Types of Landslide Inventory

As mentioned in the previous **Section 2.2**, there is some semantic variability in how landslide inventories and databases are defined, and we argue that some inventories cited in the literature may actually be better described as landslide databases comprised of different types of inventories. Indeed, the study of classifying landslide inventories into different groups is in its infancy, and we rely on a small handful of papers (Guzzetti *et al.*, 2008; Guzzetti *et al.*, 2012; Ardizzone *et al.*, 2014a; Xu, 2014; Damm and Klose, 2015) and first-hand experience working with some of these authors of these papers to outline the different taxonomy of landslide inventories in this **Section 2.3**.

Depending on the intended use of a landslide inventory map, they may be produced using different methods, at different scales and contain different levels and types of information. Landslide inventories can be broadly classified based upon:

- *Method* of inventory production (defined in **Table 2.1** and illustrated in **Figure 2.1**).
- *Age* of landslides included (defined in **Table 2.2** and illustrated in **Figure 2.2**).
- *Scale* of mapping (defined in **Table 2.3** and illustrated in **Figure 2.3**).

We will now discuss each of these three tables and figures in turn. **Table 2.1,** classification of landslide inventories by two major methods and four sub-methods of production, illustrates that *geomorphological methods* are dependent upon the geomorphological 'signature' being visible to a geomorphologist, whereas *archive methods* are dependent on landslides that humans have somehow recorded. **Figure 2.1** shows a comparison of landslide inventories produced from these geomorphological methods (a combination of sub-types) and archive methods. The inventories differ in the type of information they contain and their spatial extent: the *archive inventory* of landslides is represented as points, whereas *geomorphological inventories* are typically mapped as individual polygons for each landslide. This means that the landslide area of each polygon (A_L) can be calculated in a GIS, whereas this information may not always be contained in an archive inventory. Moreover, the *geomorphological inventory* appears to be more spatially homogeneous across the slopes of the region, whereas the *archive inventory* exhibits strong clustering around human settlements. Conversely, archive inventories may contain information that is not preserved in the geomorphological signature of a landslide, such as the triggering mechanism and impacts (Damm and Klose, 2015). **Figure 2.1** illustrates that different inventory production methods may contain different information. This will be discussed in more detail in **Section 2.4**.

Table 2.1 *Classification of Landslide Inventories by Method of Production. Production method classification and description from Guzzetti et al. (2012) with column 'examples' added by author.*

Production Method	Description	Sub-Type of Method	Examples
Geomorph-ological	Information about landslides inferred from visual assessment of field characteristics of landslides. Individual landslide extents often mapped as polygons.	Field Surveys	Cardinali *et al.* (2006); Santangelo *et al.* (2010)
		Aerial Photography	Harp and Jibson (1995); Bucknam *et al.* (2001)
		Remote Sensing	Gorum *et al.* (2011); Mondini *et al.* (2011a); Murillo-García *et al.* (2014)
		Combination of Methods	Cardinali *et al.* (2000); Lanteri and Colombo (2013)
Archive	Information about landslides gathered from secondary archive sources such as newspapers, grey literature, interviews. Generally only a point location for each landslide is given.	-	Guzzetti *et al.* (1994); Domínguez Cuesta *et al.* (1999); Guzzetti and Tonelli (2004); Kirschbaum *et al.* (2010); Petley (2012); Taylor *et al.* (2015)

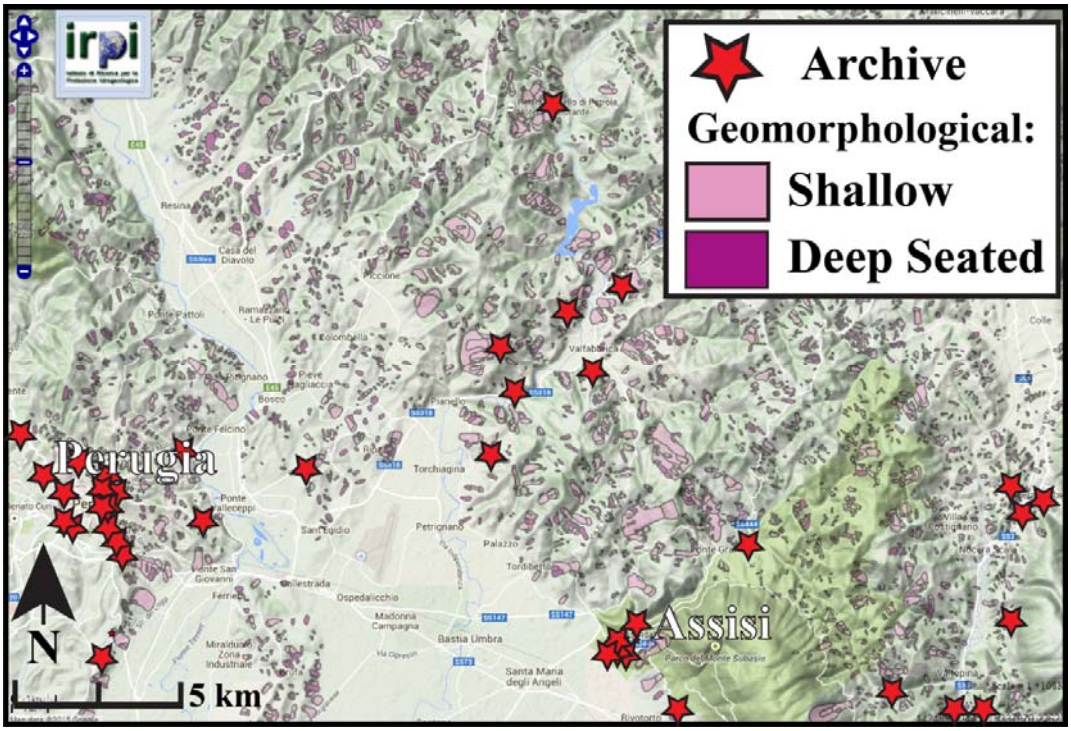


Figure 2.1 *A comparison of landslide locations using archive (red stars) and geomorphological (light and dark purple polygons) inventory methods for a region in Perugia, Central Italy. Archive method landslide locations tend to be clustered around human settlements whereas geomorphological method landslide locations are more spatially homogeneous. Map source: IRPI WebMapping (2015).*

Table 2.2 shows a classification of landslide inventory types by age (at the time of mapping) of the landslides mapped. Age may be inferred by the 'freshness' of the morphological signature (outlined later in **Table 2.4**) or the age of the source of information (e.g., the date aerial imagery was flown or the date of a newspaper article published about the landslide).

Table 2.2 *Classification of landslide inventories by age of landslides included. Age of landslides and description adapted from Guzzetti et al. (2012), with column 'examples' by author.*

Age of Landslides Included	Description	Examples
Historical (**Figure 2.2A**)	All discernible landslides are included, from ancient relict landslides to recent or active landslides. There may be some qualitative differentiation of landslide age (e.g., ancient, old, recent).	Cardinali *et al.* (2001); Duman *et al.* (2005); Dewitte *et al.* (2006)
Multi-Temporal (**Figure 2.2B**)	By exploiting multiple sets of imagery of the same location over different time periods, landslides that have occurred within various 'epochs' can be mapped and differentiated. Aside from landslides that already existed in the first set of imagery, all other landslides can be dated to the epoch of imagery in which they appear, which may be of the order of years to decades.	Carrara *et al.* (2003); Guzzetti *et al.* (2005)
Seasonal (**Figure 2.2C**)	By exploiting imagery taken a few months apart, landslides that occurred within a single season (possibly from one or two triggers) can be identified. Any landslides existing at the start of the season are excluded.	Hofmeister (2000); Fiorucci *et al.* (2011)
Event (**Figure 2.2D**)	All landslides that resulted from a single triggering event (such as heavy rainfall or an earthquake) are mapped. Any landslides that are not considered to be 'freshly' triggered are excluded.	Harp *et al.* (1981); Harp and Keefer (1990); Harp and Jibson (1995); Cardinali *et al.* (2000); Bucknam *et al.* (2001); Guzzetti *et al.* (2004a); Yagi *et al.* (2009); Gorum *et al.* (2011)

Examples of different inventories categorised by landslide age (described in **Table 2.2**) are shown in **Figure 2.2** for the Collazzone basin in Central Italy. We observe that the size of the landslides mapped is closely linked to the age of the inventory: the relict landslides mapped in **Figure 2.2A** are proportionally much larger and have less detailed mapped perimeters than those in **Figure 2.2B-D** as the morphological features of smaller landslides (approximately below the order of 1,000 m^2) will be erased from the landscape over time

(Guthrie and Evans, 2007; Bell *et al.*, 2008). The multi-temporal landslide inventory shown in **Figure 2.2B** shows a mix of landslide sizes, although there are proportionally less small landslides than shown in the seasonal and event inventories shown in **Figure 2.2C-D**. The seasonal and event inventories are visually quite similar (although the event inventory shown in **Figure 2.2D** has more landslides), as in both cases, landslides were mapped within months of being triggered and thus fewer small landslides will have been erased from the landscape. We also see proportionally fewer extremely large landslides (order of a few square kilometres) in the shorter timescale inventories (**Figures 2.2C-D**), compared to the ancient relict landslides in **Figure 2.2A** (and to a lesser extent in **Figure 2.2C**) as very large landslides occur less frequently, and possibly were triggered under different tectonic or climatic conditions (Guzzetti *et al.,* 2006a). Overall, **Figure 2.2** illustrates the importance of understanding the age of landslides mapped in an inventory when investigating the frequency-size distribution of landslides. We discuss this in more detail in **Section 2.4** and the implications of this in **Chapter 4**.

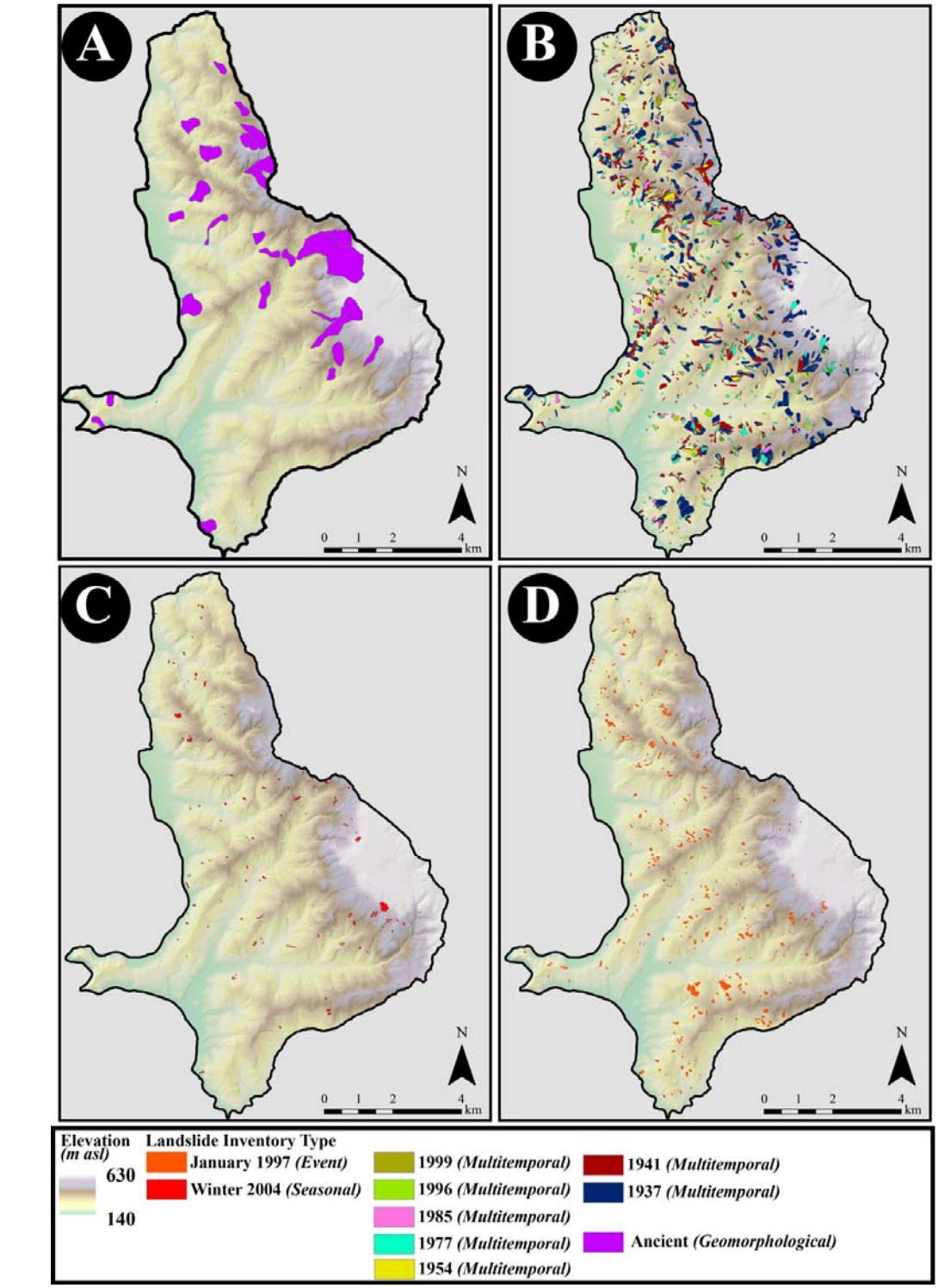


Figure 2.2 *Four different types of landslide inventory maps for the Collazzone region, Central Italy, distinguished by landslide age (**Table 2.2**).* ***(A)*** *Historical inventory of relict landslides where the exact date of occurrence is not known (Guzzetti et al., 2006a) overlaid with a 10 × 10 m² spatial resolution digital elevation model (DEM) base map (Ardizzone et al., 2007);* ***(B)*** *A multi-temporal inventory of 2,564 landslides based on analysis of aerial images taken between 1947 and 1991 and field surveys (Guzzetti et al., 2006a)* ***(C)*** *A seasonal event inventory of 70 landslides triggered in winter 2003-2004 inventoried using Airborne LiDAR (Ardizzone et al., 2007),* ***(D)*** *A triggered event inventory of 413 landslides triggered by rapid snowmelt in January 1997 inventoried from aerial photography (Cardinali et al., 2000).*

Table 2.3 shows a classification of landslides by the scale they are mapped at. The larger the scale the landslides are mapped to, the more detail there can be in the individual landslides.

Table 2.3 *Classification of landslide inventories by scale of mapping. Based on Guzzetti et al. (2012)*

Inventory Scale	Mapping Scale	Method	Purpose
Small	< 1:200,000	Archival and aerial image analysis.	Regional planning, identifying regions for large scale inventories.
Medium	1:25,000–1:200,000	Aerial image analysis (image scale 1:60,000 to 1:20,000) and field surveys.	Geomorphological and statistical studies looking at landslide patterns.
Large	> 1:25,000	Aerial image analysis (image scale > 1:20,000) and extensive field surveys.	Detailed geotechnical studies (e.g., for remedial measures at a specific site).

Figure 2.3 shows examples of landslide inventories mapped at the lower and upper end of the medium scale (**Table 2.3**) for a region in Mount Baker-Snoqualmie National Forest, Washington State, USA. As observed in **Figure 2.2**, we observe a clear dependence between scale of mapping and size and detail of landslides mapped. The larger scale landslide inventory was mapped from imagery at the 1:24,000 scale (**Figure 2.2B**) and shows many of the same landslides that are mapped in the smaller scale inventory (mapped at the 1:100,000 scale) (**Figure 2.2A**), but in more detail, and showing many additional smaller landslides that would not be discernible from smaller scale imagery. **Figure 2.3** illustrates the link between mapping scale and inventory purpose outlined in **Table 2.3**: the larger scale the map, the greater the number of small landslides included and the greater the detail. Although, this also means a greater effort is required to map and manage the data.

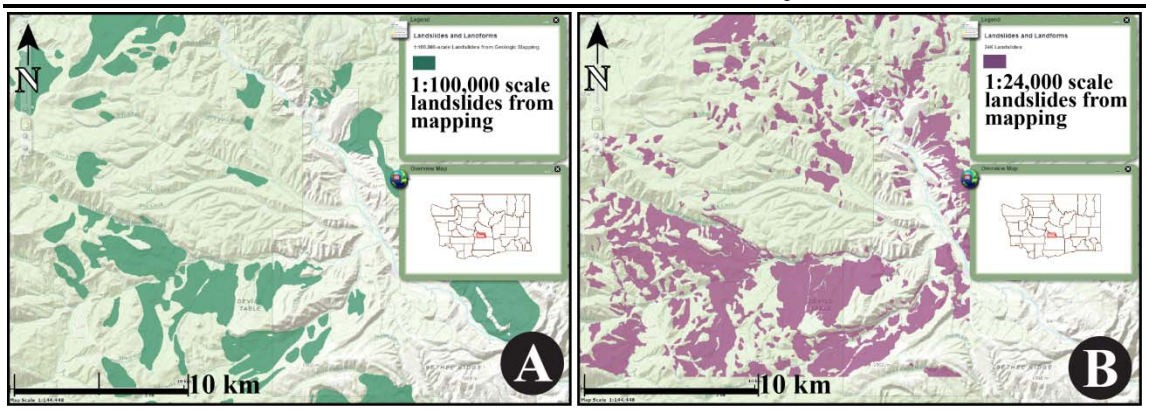


Figure 2.3 *Examples of smaller and larger medium scale landslide inventories (described in* ***Table 2.3****) for a region in Mount Baker-Snoqualmie National Forest, Washington State, USA.* ***(A)*** *Landslides mapped at 1:100,000 scale.* ***(B)*** *Landslides mapped at 1:24,000 scale. Map source: Washington Division of Geology and Earth Resources (2014).*

Because of the different motivations for creating an inventory and approaches taken (**Tables 2.1–2.3**), landslide inventories will vary in how they look (**Figures 2.1–2.3**), how they can be used and what information is contained within them. For example, landslides in an archive-based inventory may be represented as one-dimensional point locations whereas in a geomorphological inventory, the two- or three-dimensional spatial extent of the landslide may be recorded as a polygon (Carrara *et al.,* 2003). A landslide inventory geared towards understanding the human impact of landslides may be biased towards locations where there are 'landslides with consequences' and contain rich information about the timing and impact of a landslide, but less detailed geotechnical information (Guzzetti *et al.,* 2003). Whereas, an inventory geared towards scientific understanding of landscape evolution and erosional processes may aim to capture as complete a record as possible of all landslide activity across a region (Malamud *et al.,* 2004a, 2004b; Parker *et al.,* 2011).

Standards for landslide inventory production have not been widely agreed upon or taken up (Guzzetti *et al.,* 2012), and it is common to see analysis and applications of landslide inventories and databases that are arguably inappropriate for the type of data being used (e.g., Guns and Vanacker, 2014). It is important for researchers to have a clear understanding of the type of landslide inventory or database they are using and the associated uncertainties and limitations of the data. In the following **Section 2.4** we describe in detail the different types of landslide inventory available and discuss the associated uncertainties of these inventories. This sets the background for the following **Chapter 3** where we develop a method to add richness to a landslide database using an archive method. **Section 2.4** also emphasises the importance of inventory completeness,

which is important for the statistical analysis of landslide inventories reviewed in **Chapter 4** and performed in **Chapters 5** and **6**.

2.4 Methods of Landslide Inventory and Database Production

In this **Section 2.4** we discuss typical methods of landslide inventory and database production and for each section we discuss potential uncertainties involved with that particular method. These methods of inventory production are:

2.4.1 *Geomorphological Techniques*:

2.4.1.1 *Field Methods*

2.4.1.2 *Aerial Photography Methods*

2.4.1.3 *Future and Novel Remote Sensing Techniques*

2.4.1.4 *Satellite Remote Sensing*

2.4.2 *Archive Techniques*

2.4.1 Geomorphological Techniques

2.4.1.1 Field Methods

Field surveys are usually carried out in tandem with aerial photography analysis for reconnaissance and 'ground-truthing' the dataset (e.g., **Figure 2.4**), although occasionally large-scale inventories of a small region using field methods will be undertaken without accompanying imagery (e.g. Cardinali *et al.,* 2006). A team of surveyors will travel (either by foot, vehicle or helicopter/plane) around a region to identify landslides. When a landslide is located, its location, extent and characteristics can be mapped using tools/methods such as tachometry, transit and tape, electronic distance measurement, global positioning systems (GPS), ground-based photography, local interviews and total station equipment approaches (Clayton *et al.,* 1982; Keaton and de Graff, 1996).

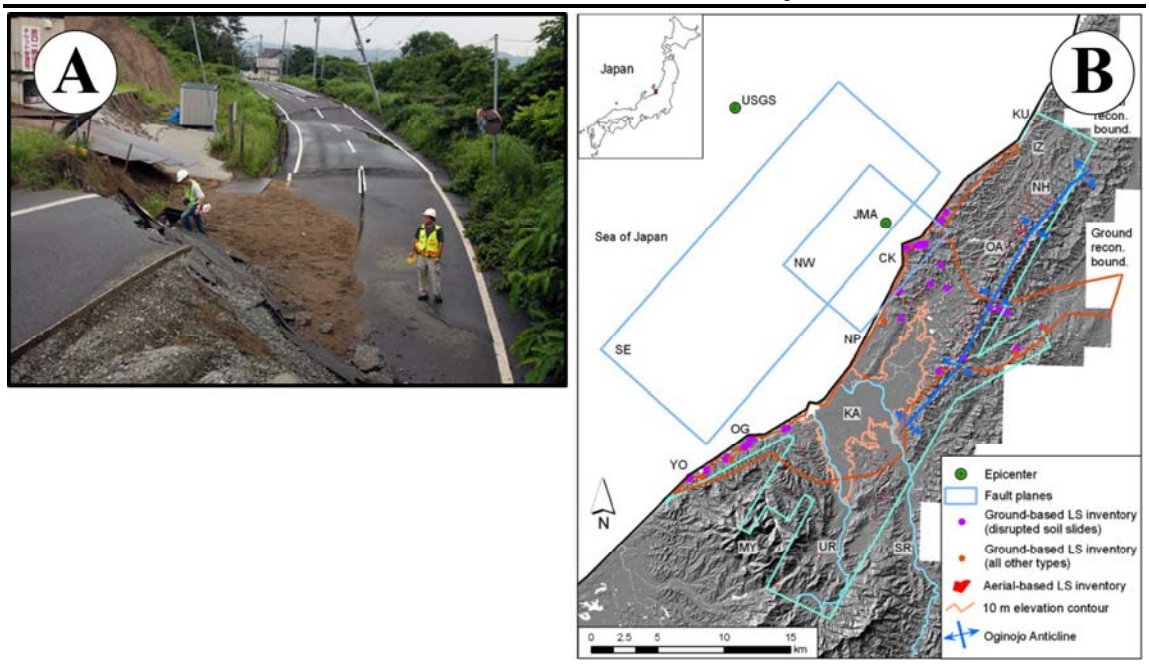


Figure 2.4 *USGS scientists performing field inventory of landslides triggered by the moment magnitude (M_W) = 6.6 Nigata Chuetsu Oki Earthquake in Japan, 2007.* ***(A)*** *Photos of the scientists recording observations of landslides and damage using cameras, global positioning systems (GPS) Image source: Kayen (2008),* ***(B)*** *Resulting field inventory of landslides triggered by the Earthquake (purple and orange dots). Map source: Collins et al. (2012).*

Uncertainties with Field Methods

Field methods cannot be considered as a uniform survey across a region for a number of reasons:

- *Accessibility.* There is the issue of what can be accessed, either due to land ownership issues or physically reaching a site (e.g., in a region with few roads or dangerous terrain). This is illustrated in **Figure 2.4B** where landslides mapped by driving around in the field tend to be more clustered than those mapped from aerial photography, presumably due to issues of vehicular access.
- *Time constraints.* Due to constraints of time, the geomorphologists will not map very small landslides (with 'small' landslide area (A_L) defined by the particular mapping being done, e.g., $A_L < 5$ m^2) so for each small landslide seen, an 'in-the-field' decision must be made whether or not to map the landslide, which may vary depending on the expert opinion (M. Rossi, pers. comm., January 2012). Arguably, these landslides are negligible as they are mostly smaller than those identified through aerial photo analysis, but there is an issue of sampling uniformity across the study region.
- *Overall perspective difficulties.* At the other end of the scale, it is difficult to recognise very large landslides and slow moving failures (i.e. entire slope faces, approximately

A_L > 1 km^2) as the perspective requires some distance to be recognisable (Guzzetti *et al.*, 2012).

2.4.1.2 Aerial Photography

Aerial photography is perhaps the most commonly used method of landslide inventory production and has been so since the middle of the 20th century (Keefer, 2002, Guzzetti *et al.*, 2012). Most typically, this is performed using relatively high resolution (cell size < 3 m × 3 m) stereoscopic imagery (van Westen *et al.*, 2008) where a pair of photographs focused on the same location but taken from different angles are combined using a stereoscope instrument (**Figure 2.5A**) to give the viewer the illusion of image depth. This pseudo-3D image aids the user in identifying the location and outline of landslides, which are typically mapped by hand using transparency overlays and later digitised to produce a landslide inventory map, as shown in **Figure 2.5B**.

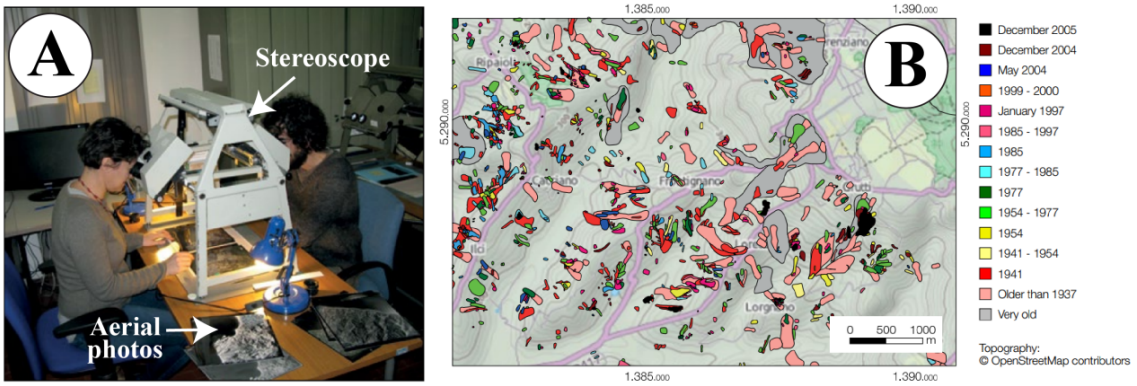


Figure 2.5 (A) *Example of geomorphologists using a stereoscope to produce an inventory of landslides from stereoscopic aerial imagery and* ***(B)*** *the resulting multi-temporal landslide inventory created from several sets of aerial imagery for the Umbria Region, Central Italy. Figure source: LAMPRE (2014).*

Although digital stereoscopy and airborne LiDAR systems (discussed in **Section 2.4.1.3**) are becoming more widely used, often this remains an analogue process using a stereoscope, physical photographs and a transparency overlay onto which the landslides are drawn by hand. These transparencies are then mapped onto a topographic map using neighbouring features as landmarks, which can then be digitized into a GIS. This method is time consuming and labour intensive: Galli *et al.* (2008) found that interpreters analysed 8 km^2 – 470 km^2 of aerial imagery per person per month, depending on the scale of imagery and the number of landslides, thus it can take a small team (2–3 persons) months to years to produce large landslide inventories. Although time consuming, this method has the benefits of being relatively low cost and allowing systematic coverage of relatively large expanses of land.

The pseudo-3D image discussed above aids the operator to interpret the morphological features that make up the 'signature' of a landslide. **Table 2.4** highlights the wide range of tonal, textural, morphological and contextual features that can visually signify the presence of a landslide in aerial photography, which will vary depending on location and types of landslide observed, and can be very subtle to the untrained eye. Typically, the ability to recognise these features is learnt empirically over a number of years practice (Guzzetti *et al.*, 2008), and consequently can vary between geomorphologists. This is discussed further in the following section.

Table 2.4 *Characteristic features of landslides used to locate and measure landslides in aerial image analysis (Soeters and van Western, 1996 with minor additional definitions and clarifications by author).*

Terrain Features	Relation to Slope Instability	Photographic Characteristics
Morphology		
Concave/convex slope features	Landslide crown and associated deposit	Concave/convex anomalies in stereo model
Step-like morphology	Retrogressive Sliding	Step-like appearance of slope
Semi-circular back-scarp and steps	Head part of slide with outcrop of failure plane	Light-toned scarp (cliff), associated with small, slightly curved lineaments (linear surface concavities aligned with faults)
Back-tilting of slope facets	Rotational movement of slide blocks	Oval or elongated depressions with imperfect drainage conditions
Hummocky and irregular slope morphology	Micro-relief associated with shallow movements or small retrogressive slide blocks	Coarse surface texture, contrasting with smooth surroundings
In-filled valleys with slight convex bottom where V-shaped valleys are normal.	Mass movement deposit of rock/debris/mud flow	Anomaly in valley morphology, often with lobate form and flow pattern on the body.
Vegetation		
Vegetation clearances on steep scarps, coinciding with morphological steps	Absence of vegetation on head scarp or steps	Light toned elongated areas at crown
Irregular linear clearances along slope	Slip surface of translational slide and/or path cleared by a flow.	Denuded (exposed) areas showing light tones, often with linear pattern in direction of movement
Disrupted, disordered and partly dead vegetation	Slide blocks and differential movements in body	Irregular or mottled grey tones
Differential vegetation associated with changing drainage conditions	Stagnated drainage on back-tilting blocks and seepage at toe	Tonal differences displayed in pattern associated with morphological anomalies in stereo model

Terrain Features	Relation to Slope Instability	Photographic Characteristics
Drainage		
Areas with stagnated drainage	Landslide niche, back-tilting landslide blocks and hummocky internal relief on landslide body	Tonal differences with darker tones associated with wetter areas
Excessively drained areas	Out-bulging landslide body (with differential vegetation and some soil erosion)	Light toned zones in association with convex relief forms
Seepage and spring levels	Springs along frontal lobe and at places where failure plane outcrops	Dark patches sometimes in slightly curved pattern and enhanced by differential vegetation
Interruption of drainage lines	Drainage anomaly caused by head scarp	Drainage line abruptly broken off on slope by steeper relief
Anomalous drainage pattern	Streams curving around frontal lobe or streams on both sides of body	Curved drainage pattern upstream with sedimentation of meandering in valley
Deposition in stream channels	Deposition of flow material into river channels	Lighter tones in river channel compared with surrounding areas

Issues with 'Unseen' Landslides and Subjective Interpretation in Aerial Photography

For 'fresh' landslides, identification of them from aerial imagery is relatively simple where there is clear and abrupt change in vegetation or surface texture. For older landslides erosion, revegetation, anthropic activity and further landsliding may obscure the boundary of the landslide, and thus require more inference from the geomorphologist (Malamud *et al.,* 2004a). Moreover Brardinoni *et al.* (2003) estimated, when comparing inventories of landslides produced from aerial photo analysis and field investigation, that up to 85 % of individual landslides in certain basins were not visible from aerial photography because they were obscured by forest (although there was significant variability depending on the basin examined). Harp and Jibson (1995) also noted, in a study of 11,111 landslides triggered by the Northridge Earthquake (USA), that the resolvable size of landslides visible on sunlit slopes was higher than those on shaded slopes, resulting in irregular sampling across a region. This issue of only partially-resolvable landslides is confounded by the fact that there are not agreed-upon standards of landslide identification from aerial photography for geomorphologists, and this remains somewhat of a subjective 'art form' that is learnt over time and varies from person to person (Guzzetti *et al.,* 2012; Xu, 2014). Indeed, Ardizzone

et al. (1999) found that when three independent experts produced an historical inventory map from the same image, there were often differences between the boundaries of mapped landslides and the positional mismatch between boundaries classified as 'certainly unstable' by the different experts was up to 80 %. Certainly, the lack of clear standards and how training is undertaken can result in discrepancies in the number and detail of landslides mapped from aerial photography.

Issues with Data Management and Transfer of Landslides Mapped from Aerial Photography

The previous section outlined some of the uncertainties relating to whether or not landslides are initially seen and mapped by a geomorphologist. However, there are also issues relating to the physical process of creating, transferring and managing spatial data that can introduce errors and biases into an inventory map.

- *Initial tracing of the outline of the landslide.* When tracing, a small deviation (1–2 mm) from the 'true' landslide boundary will result in 10–20 m error on the ground for large scale inventories (at 1:10,000 scale), resulting in inaccuracies in area estimation which will be greatest for smaller landslides (Malamud *et al.,* 2004a).
- *Traced outlines are then transferred to a topographic map.* These are based on the features of that map (such as drainage divides, settlements, roads and land cover). Positioning errors can arise if the topographic map is incorrect or has few features (Malamud *et al.,* 2004a).
- *Digitising and converting the outlines to vector format.* This allows calculating the landslide area of the polygon. As airborne stereoscopic imagery has a conical map projection and commonly used digital formats have a cylindrical projection, neither is representative of the true area (Longley *et al.,* 2010). Moreover, as the majority of landslides occur on sloped terrain where the camera is at an oblique angle, the dimension of the landslide parallel to the slope may be truncated, thus the landslide area calculated in a GIS will be subject to error (Fernández *et al.,* 2006). The user must consider whether the above distortions have been corrected for, which is not always stated in the metadata.
- *Recording of metadata.* It is not always clear in the metadata what areas are being mapped; sometimes confusion can arise over whether the head and crown are mapped separately, or sometimes more than one landslide area is aggregated into

one polygon (where the boundaries between landslides are unclear) (Malamud *et al.,* 2004a).

Using a digital system of digital or scanned stereoscopic images and directly digitizing landslide vectors within a computer circumvents some of the issues relating to digitization discussed above, and there are a small number of examples in the literature that have adopted this process (e.g., Gao and Maro 2010; Fiorucci *et al.,* 2011). However, this digital system has been applied more widely using satellite remote sensing imagery (discussed in **Section 2.2.1.3**), most likely because of the increase in availability of high resolution satellite stereoscopic imagery in previous years (sidestepping the need to commission aerial imagery). The lack of uptake across the board to date might also be attributed to the cost of purchasing equipment and retraining geomorphologists in these new systems (Guzzetti *et al.,* 2012).

2.4.1.3 Future and Novel Airborne Imagery Landslide Inventory Techniques

LiDAR and SAR

Although not yet widely applied, active remote sensing techniques such as Light Detection And Ranging (LiDAR) (**Figure 2.6A**) and to a lesser extent Synthetic Aperture Radar (SAR) show promise for future landslide inventory production (Metternicht *et al.,* 2005; Ardizzone *et al.*, 2007; van Westen *et al.,* 2008; Jaboyedoff *et al.,* 2012). These sensors are used on both aerial and satellite platforms. These techniques are not dependent upon energy reflected or emitted by the Earth's surface but instead emit their own source of radiation and measure the reflectance returned back to the sensor (Jensen, 2000). By measuring the time taken for the energy to be reflected back to the sensor from each point on the ground, a digital elevation model of the surface can be created (Jaboyedoff *et al.,* 2012) (**Figure 2.6B**). Landslides can then be manually identified from their morphometric features (**Table 2.4**) or semi-automatically detected by looking at the difference between DEMs taken at different time periods (Scaioni *et al.*, 2014).

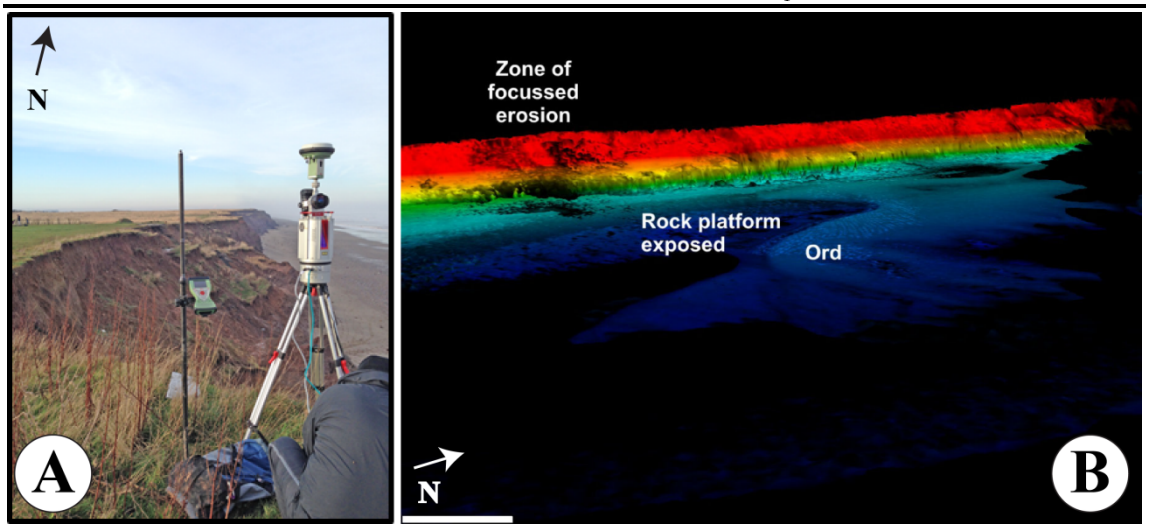


Figure 2.6 *Example of Ground based LiDAR equipment being used to survey cliff collapse in Aldbrough, United Kingdom.* ***(A)*** *LiDAR equipment on the top of the cliffs in January 2014. Image: Faith Taylor (2014)* ***(B)*** *Resulting 3D model of cliffs produced from placing the LiDAR equipment on the beach slightly south of the location shown in A. Image: BGS (2015).*

By using active remote sensing methods, it may be possible to measure ground surface deformations under conditions where passive remote sensing techniques would not be applicable. For example, both LiDAR and SAR work under low-light levels (Liew, 2001), LiDAR can image the ground surface through vegetation canopies (McKean and Roering, 2004) and SAR can image through heavy cloud cover (Singhroy, 1995); all of these are limitations of passive optical remote sensing. Airborne LiDAR has been applied successfully for landslide inventory production in a few examples: Ardizzone *et al.* (2007) compared a field-based and LiDAR based triggered event inventory for a 90 km^2 region in Italy and found 27 % increase in the number of landslides mapped (N_{LS}), 39 % less total landslide area (A_L) (attributed to higher precision measurement of A_L) and that the spatial precision of LiDAR mapped landslides was better than field techniques. Chigira *et al.* (2004) created an inventory of $N_{LS} > 1{,}000$ rainfall triggered landslides in Japan and identified additional landslides that had previously been obscured from vegetation using traditional techniques. Nonetheless Jabadoyeff *et al.* (2012) note issues such as areas being shadowed (and thus not mapped) by rough terrain and issues with handling and merging the large volumes of data created.

Examples of SAR in the literature are more focused upon remote sensing of the characteristics of individual landslides using differencing between multiple digital elevation models (DEMs) (referred to as Differential Interferometric SAR, DInSAR) rather than the production of full inventories (Metternicht *et al.,* 2005; Guzzetti *et al.*, 2012). This is most likely the result of issues of temporal and spatial decorrelation (i.e., difficulty in identifying

the same location over multiple sets of imagery) caused by changes in the reflectance characteristics of certain ground cover types; particularly vegetated surfaces. Nonetheless, authors such as Farina *et al.* (2006) demonstrated that DInSAR techniques can complement inventories created from traditional techniques by identifying higher resolution boundaries of landslides and a small number of additional landslides normally obscured by vegetation. Certainly, as these techniques are developed and more widely applied, they could produce landslide inventories with considerably more information and detail on individual landslides than traditional techniques.

Unmanned Aerial Vehicles (UAVs or 'Drones')

In the past few years, the use of unmanned aerial vehicles (UAVs, commonly referred to as Drones) to monitor individual large landslides and inventory small areas has also showed promise. One method used to create digital elevation models (DEMs) of slopes is Structure from Motion (SfM), whereby a standard digital camera is attached to the UAV and flown across a region. Much like traditional stereoscopy, the camera captures multiple photographs of the same location from different angles, which can then be triangulated to build a three dimensional model (Fonstad *et al.*, 2013). Landslides can then be identified from morphological characteristics or by applying change detection between multiple DEMs. Several examples in the literature exist using this technology to monitor individual landslides (e.g., Niethammer *et al.,* 2012; Wen *et al.,* 2011), and a small number of papers/grey literature (e.g., Rau *et al.,* 2011) and research groups (e.g., the University of Northumbria (The Grand Alpine Tour, 2014)) working on inventory production using UAVs.

UAVs have the benefit of being relatively low cost, flying at a low height (and thus having a high spatial resolution) and being portable/flexible so can rapidly be deployed following a landslide event. As this technology is relatively new, it is not exactly clear what the associated uncertainties related to inventory production will be, although Fonstad *et al.* (2013) highlighted issues with interoperability with traditional software packages and Niethammer *et al.* (2012) identified problems with 'missing' sections of DEM where the land was obscured by vegetation or topography and issues with time taken to process and georectify the data. One interesting consideration is the 'democratization' of landslide inventories; as UAV technology is cheap, increasingly accessible, requires little training and DEMs can be produced using open-source software, there may be an abundance of data created by 'citizen scientists' (Jordan *et al.,* 2011). Without clear guidelines and standards

(even for traditional methods of inventory production (Guzzetti *et al.*, 2012)), this may raise interesting questions about data availability versus quality (Dickinson *et al.,* 2010).

Seismic Arrays

Ekström and Stark (2013) recently demonstrated that arrays of seismic sensors normally used for detecting and monitoring earthquakes can be used to identify the timing, location and characteristics of large landslides. These landslides exhibit a characteristic low-frequency signal caused by the loading and unloading of material. Although as of yet only shown for large landslides, this method has some complimentary features to traditional geomorphological methods as the seismic signal includes information that is not recorded in the geomorphological signature of a landslide such as the time of occurrence and the duration and nature of the movement. Moreover, seismic sensors were shown to detect landslides that are very remote from human activity and would otherwise have possibly gone undetected using geomorphological or archive techniques.

2.4.1.4 Remote Sensing

Here, remotely-located sensors are defined as satellite platforms which are orbiting the Earth housing various types of sensor which measure energy reflected from the Earth's surface (although broader definitions would include instrumentation such as airborne photography, binoculars, etc.). Landslides can be identified from their reflectance characteristics and mapped (example shown in **Figure 2.7**), using methods such as:

- *Manual identification* using techniques summarised in **Section 2.5**.
- *Change detection* using images before and after a landslide-triggering event (Mondini *et al.,* 2011).
- *Semi-automated techniques* based upon the spectral signature (i.e., reflectance in different wavebands) of a landslide (Daneels *et al.,* 2007).
- *Active remote sensing techniques* such as InSAR and LiDAR (described above) which create a digital elevation model.
- *A combination* of methods (e.g., Mondini and Chang, 2014).

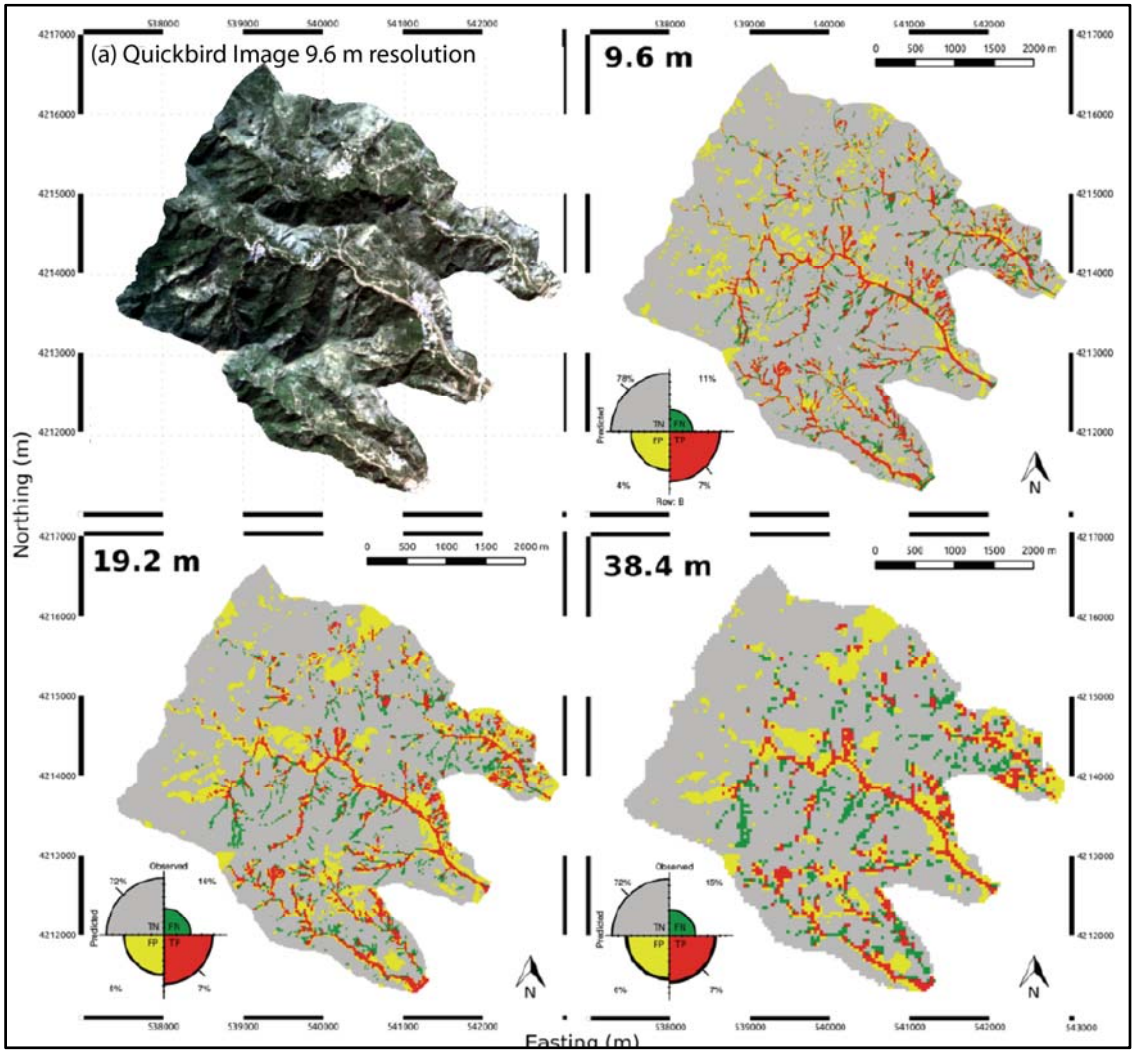


Figure 2.7 *Example of landslide inventory created using satellite remote sensing imagery degraded to three different spatial resolutions for the Messina Region, Italy.* ***(A)*** *Quickbird satellite imagery at 9.6 m spatial resolution. Inventories created from imagery at* ***(B)*** *9.6 m;* ***(C)*** *19.2 m and* ***(D)*** *38.4 m spatial resolution. As the resolution of the imagery decreases, the number of landslides mapped decreases, as does the detail they can be mapped to. True negatives (TN) are shown in grey and represent zones that are not landslide and were classified as not landslide in the remote sensing imagery. True positives (TP) are red and show zones that are landslides and were classified as landslide in the remote sensing imagery. False positives (FP) are shown in yellow and are areas that were identified as landslide in the remote sensing imagery but were not truly landslides. False negatives (FN) are shown in green and represent areas where a landslide had occurred but was not detected in the remote sensing imagery. Figure source: Mota et al. (2014).*

An extensive discussion and review of papers that use remote sensing for landslide inventory compilation is out of the scope of this chapter, due to their increasingly larger numbers in recent years and extensive reviews have been compiled by others such as Scaioni *et al.* (2014). We instead focus on describing some of the overarching uncertainties related to all of the above remote sensing methods which can result in variability in the number, size and detail of landslides mapped.

Issues and Uncertainties with Landslide Inventory Production Using Remote Sensing

These methods may allow precise and rapid inventorying over wide regions, but are also subject to errors of omission and commission:

- *Errors of omission*. These occur where landslides are not mapped as they do not follow the rules of classification as a landslide, such as a translational movement where vegetation remains intact or landslide size is below the spatial resolution of the sensor (Nichol and Wong, 2005).
- *Errors of commission*. These occur where land cover types with a similar spectral signature to those of landslides (such as freshly ploughed fields or roads) are incorrectly classified as landslides, thus adding erroneous landslide areas to the inventory (Nichol and Wong, 2007).

The *number of errors of omission and commission* can vary greatly depending on the land cover type, imagery resolution and techniques or algorithms used to identify landslides (Scaioni *et al.,* 2014). Presently, lower numbers of errors of omission/commission are found in forested regions with little human activity, where landslides appear as an abrupt and distinct change in land cover (Borghius *et al.,* 2007). In regions where the surrounding land has little textural or tonal difference to the landslide, a greater number of errors are observed (Nichol and Wong, 2005), or for large deep seated landslides where the earth surface is not significantly disrupted by the movement, and thus there is no change in texture or tone (Guzzetti *et al.,* 2012).

The ability to produce a landslide inventory from satellite remote sensing imagery is also affected by the *spatial resolution* of the platform. **Figure 2.7** demonstrates three triggered event landslide inventories created using increasingly coarse spatial resolution Quickbird satellite imagery. As the pixel size of the imagery increases, the number of errors of omission increases as more landslides are below the detectable resolution of the imagery. Also, in the coarser resolution imagery, the detail to which a landslide perimeter and area can be mapped declines due to the more pixelated nature of the imagery, resulting in a bias towards larger landslide areas.

As an example of the variability involved when different scientists estimates of the number of landslides (N_{LS}) triggered by the same event, the 2010 Haiti earthquake (M_w = 7.9) number of estimated landslides triggered ranged from N_{LS} > 4,492 landslides (Gorum *et al.,*

2013) to N_{LS} > 7,000 landslides (Harp *et al.*, 2013) to N_{LS} = 30,028 landslides (Xu *et al.*, 2014) using the same set of high resolution imagery but different scientists and approaches. Similar variations in quoted number of landslides triggered by the 2008 Wenchuan Earthquake (M_W = 7.9) have been noted, for example, Huang and Li (2009) inventoried N_{LS} = 11,308 landslides and estimated the total number of landslides was N_{LS} > 50,000; Yin *et al.* (2009) estimated N_{LS} > 15,000 landslides; Dai *et al.* (2011) inventoried N_{LS} > 56,000 landslides and Xu (2014) inventoried N_{LS} = 197,481 landslides.

In the 2010 Haiti earthquake inventories compiled by the different authors mentioned above, there are differences in the areal extent of the region considered to be affected, but when the density of landslides is compared, this does fully not account for discrepancies in results. In the Wenchuan earthquake inventories, there are differences in the types of landslide mapped: Yin *et al.* (2009) focused on landslides that had affected human activity whereas Dai *et al.* (2011) inventoried landslides with areas A_L > 500 m^2 and the methodology of Huang and Li (2009) is not explicitly clear.

These examples illustrate that the communication of what landslides are being counted and how the extent of the affected area is defined (Crozier, 2005) can have significant impacts on the number and spatial distribution of landslides included in an inventory. Although each inventory may be suitable for its original purpose, if the methods and motivations for production are not clearly stated, this can result in misinterpretation and misuse of inventories in later analysis. Nonetheless, as higher spatial and spectral resolution imagery becomes increasingly available, it is likely that more robust methodologies will evolve and be used more widely.

The geomorphological methods discussed in **Section 2.3.1** pertain to mapping landslides that have left some discernible signature on the landscape. This means that landslides mapped using these methods are generally fairly recent (within the epoch of aerial photography and remote sensing) or old landslides that are very large, and thus have not been erased from the landscape over time. These methods are also limited by the number of skilled human-hours available and data that are accessible. In the next **Section 2.4.2** we describe archive methods of creating landslide inventories which can help address some of these issues with geomorphological methods.

2.4.2 Archive Methods of Landslide Inventory Production

Archive methods of producing landslide inventories involve searching archives of human generated content such as newspapers, reports, diaries, maps, photographs, theses, emergency declarations, grey literature, interviews and more recently, online material such as social media and citizen science reporting (e.g., Guzzetti *et al.*, 1994; Glade and Crozier, 1996; Domínguez Cuesta *et al.*, 1999; Guzzetti and Tonelli, 2004; Devoli *et al.*, 2007; Petley, 2012; Pennington *et al.*, 2015). This approach can be used as a single method (e.g., Kirschbaum *et al.,* 2010; Petley, 2012), or more commonly in conjunction with geomorphological methods to supplement existing landslide inventories (e.g., Pennington *et al.,* 2015). Inventories may be supplemented by (i) increasing the spatial extent of records (e.g., Devoli *et al.,* 2007), (ii) extending the records further back in time (Guzzetti *et al.,* 1994) and (iii) adding richer detail to existing records (e.g., Taylor *et al.,* 2015, **Chapter 3** of this thesis).

This primarily desk-based approach can be considered relatively low-cost and arguably requiring less skill in landslide identification methods, and has been applied successfully in a number of papers to considerably increase the number of records in landslide databases (Taylor *et al.*, 2015). Perhaps the most cited example is the Italian AVI (Aree Vulnerate Italianne [Areas Affected by Landslides and Floods]) project (available online at CNR–GNDCI, 2012), containing records of > 32,000 landslide and > 29,000 flood events, going back 1,000 years, but with most recorded between 1900 and early 2000, of which ~78 % of the information comes from newspaper reports (Guzzetti *et al.,* 1994; Guzzetti and Tonelli, 2004). More recently, the growing capacity to search freely available digital archives of global newspaper reports and online sources has prompted the construction of the Durham Fatal Landslide Database, which includes fatal rainfall triggered landslide events occurring across the globe since 2004. For the seven year period, 2004−2010, the database includes 2,620 fatal landslides, which resulted in 32,322 fatalities (Petley, 2012). Other examples of landslide databases using newspaper articles as a source of information include Domínguez Cuesta *et al.* (1999) in the North of Spain; Glade and Crozier (1996) in New Zealand; Devoli *et al.* (2007) in Nicaragua; and Taylor *et al.* (2015) in Great Britain.

Issues and Uncertainties of Archive Methods of Landslide Inventory Production

The type of landslide information available from archive materials generally differs to those using geomorphological methods. For example, Taylor *et al.* (2015) noted that newspaper

articles from Great Britain were a good source of information about the timing and impact of landslides, but rarely contained detailed geotechnical information such as the type and size of landslide. More commonly, landslides are represented as a point rather than a polygon area, limiting the types of analysis that can be performed.

There are also spatial and temporal biases in archive materials as a proxy for information about landslides, such as an overemphasis on events with human impact (Carrara *et al.*, 2003), increased media interest following a number of events (Pennington and Harrison, 2013), a focus on high magnitude events or underreporting of low magnitude events (Guzzetti and Tonelli, 2004) and scientific correctness of information (Ibsen and Brunsden, 1996).

2.5 Landslide Database and Inventory 'Completeness' and Implications for Research Performed in this Thesis

In **Section 2.4** we described the main uncertainties encountered when creating a catalogue of landslides (inventory or database). Throughout the **Section 2.4**, we emphasised that there is a strong link between the methods used and the number, size and detail to which landslides are mapped, meaning that it is relatively rare for an inventory or database to be considered a 'complete' record of landslides, certainly over all time, but even in a single triggered landslide event. For example, in a review of 24 national landslide databases from European countries, van Den Eeckhaut and Hervás (2012) found that 14/21 database managers estimated their databases to be < 50 % complete in terms of recording all landslides that have occurred in that country (3 did not respond). Moreover, 5/24 database managers stated that landslide records did not contain information about landslide area or volume and 7/24 database managers stated that < 50 % of landslide records contained information about impact (van Den Eeckhaut and Hervás, 2012), highlighting the issue that even if a database can be considered approaching 'completeness' in terms of number of landslides, it may not contain all fields of information such as landslide size and impact. In the following **Chapter 3** we examine this issue in more detail with regard to the *National Landslide Database of Great Britain*, where we look at ways to (i) add additional records of landslides but also (ii) ways to add more information about each landslide and potentially restructure how landslide impact data is recorded.

In **Chapters 4–7** of this thesis, we review and characterise the statistical behaviour of landslide size (**Chapter 4**), shape (**Chapter 5**) and spatial clustering (**Chapter 6**) of landslides and use this statistical behaviour to create synthetic triggered landslide event

inventory maps in the LRIM (**Chapter 7**). In order to understand this statistical behaviour, we must work with triggered landslide event inventories that represent the full range of landslide size, shape and spatial location. This means we must look at inventories that can be considered relatively 'complete' and lacking bias (Malamud *et al.,* 2004a). Thus, we focus on triggered landslide event inventories where we can assume that (i) all landslides mapped resulted from a single trigger and (ii) the majority of landslides that were triggered in that event are mapped. This means we can approximately assume that an inventory as a statistical 'population' of landslides. This assumption is dependent on using triggered landslide inventories where the uncertainties described in **Section 2.4** are minimised, and the inventory is well documented. Harp *et al.* (2011) outline seven criteria for the imagery and methods used to produce an earthquake triggered landslide inventory to allow robust analysis (**Table 2.5**), which we argue is also applicable for other types of triggered landslide event inventories.

Table 2.5 *Criteria for the imagery and methods used to create a substantially complete, statistically robust triggered landslide event inventory. Quoted verbatim from Harp et al. (2011).*

Number	Criteria
1	The imagery must be continuous and span the entire landslide distribution.
2	The imagery must have a resolution that allows identification of individual landslides as small as a few meters across.
3	The imagery must have stereo coverage or be able to be draped over a digital elevation model to obtain a stereo-like perspective view.
4	The imagery (as cloud-free as possible) must be acquired as soon as possible after the earthquake to capture the initial aspects of the landslides and the terrain or infrastructure that they affect.
5	The landslides must be defined as polygons in a GIS program, either as a single polygon representing the entire landslide or, as two or more polygons that define the landslide source and the landslide deposit.
6	The landslide polygons must be plotted on a topographic map or GIS layer that is registered to a topographic map or geo-registered image.
7	The entire population of landslides triggered by the earthquake must be mapped and their margins digitized. All landslides that exceed the minimum resolution of the imagery must be mapped and plotted so that a complete landslide distribution can be obtained.

Because of methodological issues (described in **Section 2.4**), a lack of clear standards and often-poor metadata that does not fully describe a dataset, many inventories do not meet the seven criteria given in **Table 2.5**. Moreover, production of these inventories is a large task (of the order of months to years of work) that may not be financially justifiable for research scientists or government agencies to produce and share when less stringent landslide inventories may suffice for their intended applications. Indeed, Harp *et al.* (2011) identified just ten earthquake triggered inventories globally that met their criteria, of which

only one is freely publically available as a spatially referenced dataset; in our experience the same lack of substantially complete 'quality' inventories is observed for other types of triggering events (e.g., heavy rain).

Major issues relevant to analyses performed later in this thesis are the following:

(i) *Under-sampling* of small landslides that are either below visible resolution or quickly erased from the landscape in comparison to large landslides.

(ii) *Not all landslides being associated with the triggering event* (i.e., some landslides that occurred before the trigger may be included, some landslides or study areas affected by the trigger may not have been mapped and difficulties in defining a 'cut off' after which any new landslides cannot be considered to have resulted from the specific trigger).

(iii) *Errors with mapping* such as the location of landslides and the detail to which the outline is mapped.

(iv) *Lack of clarity about separation of landslide polygons*, particularly separation of source area and deposits (particularly for debris flows) and separation of landslide complexes.

(v) *Lack of information about landslide impacts*. Where available, this information is often recorded and stored differently by a variety of public and private organisations involved in remediating landslide impacts.

As a result, analyses in **Chapters 4–7** are limited to a small number of suitable and available inventories, which are described in **Chapters 4** and **5**. It is hoped that as research groups work to define standards for inventory production (e.g., Ardizzone *et al.*, 2014a), as very high-resolution satellite imagery becomes increasingly available, and as many scientists move towards open source data, that more of these substantially complete triggered landslide event inventories become available in the coming decade.

2.6 Summary and Conclusions

In this **Chapter 2**, we have found:

- There is a *lack of clarity in the existing literature regarding what a landslide inventory or database is*, how they can be classified into types, and what types of analyses are appropriate for different types of inventory

  - We define a landslide database as a collection of records of landslides, which could be a combination of several inventories and records of individual landslides.
  - We define a landslide inventory as a catalogue that aims to record a complete 'set' of landslides.

- The *major methods of landslide inventory and database production are*: field techniques, aerial photo interpretation, novel ground and airborne remote sensing techniques, satellite remote sensing and archive methods.
- There is a *strong link between what methods are used to produce an inventory/database and completeness and detail of that landslide inventory or database*:
  - **Field techniques** tend to be biased towards locations that can be accessed by foot or vehicle.
  - **Aerial photo interpretation** is perhaps the most commonly used technique but is somewhat subjective due to the lack of agreed upon standards and dependence on the geomorphologist being able to interpret the geomorphological signature of a landslide.
  - **Novel techniques such as LiDAR and UAVs** show promise for the rapid production of digital elevation models from which landslides can be interpreted, although there has not been a clear debate within the literature of the limitations and uncertainties of these techniques, and few inventories produced this way are publically available.
  - **Satellite remote sensing techniques** also show promise for the rapid inventorying of large areas affected by landslides, although are subject to errors of omission if landslides do not have a strong spectral signature or errors of commission if other features have a similar spectral signature to landslides.
  - **Archive techniques** provide a useful way to supplement databases and inventories in cases where landslides do not have a geomorphological signature (either because they have been erased from the landscape over time or because there is no access to geomorphological techniques), although they tend to be biased towards landslides that have affected humans.
- It is *difficult to achieve or approach inventory/database completeness* from any method, in terms of the number of landslides recorded or the information available for each of those landslides.

- Broadly, *smaller landslide areas tend to be under represented in landslide inventories and databases* as they are either (i) less visible than larger landslides or (ii) more rapidly erased from the landscape via erosional and anthropogenic processes before they can be catalogued.
- *Landslide inventory completeness* in terms of number of landslides mapped and the representativeness of the sample has *implications for the statistical analysis* reviewed and performed in **Chapters 4–7** of this thesis.
- We explain that a *substantially complete triggered landslide event inventory is the most statistically representative sample of overall landslide behaviour* in terms of the number, shape, size and spatial distribution of landslides. We describe a criteria developed by Harp *et al.* (2011) for a substantially complete triggered landslide event inventory, which we use to select inventories for analysis in **Chapters 5–7**.

In the following **Chapter 3** of this thesis, we present a method to supplement a database of landslides with more records and more information about landslides.

Chapter 3: Adding Richness to the National Landslide Database of Great Britain

Chapter Summary

In **Chapter 2**, we highlighted how methods used to catalogue landslides mean that most landslide databases and inventories are incomplete. In this **Chapter 3**, we present a method to supplement existing records of landslides in Great Britain by searching an electronic archive of regional newspapers. This method does not aim to 'complete' the database, but create a richer set of data with which to understand landslide impact. In Great Britain, the British Geological Survey (BGS) is responsible for updating and maintaining records of landslide events and their impacts in the National Landslide Database (NLD). The NLD contains records of more than 16,500 landslide events in Great Britain. Data sources for the NLD include field surveys, academic articles, grey literature, news, public reports and, since 2012, social media. We aim to supplement the richness of the NLD by (i) identifying additional landslide events, (ii) acting as an additional source of confirmation of events existing in the NLD and (iii) adding more detail to existing database entries. This is done by systematically searching the Nexis UK digital archive of 568 regional newspapers published in the UK. We construct a robust Boolean search criterion by experimenting with landslide terminology for four training periods. We then apply this search to all articles published in 2006 and 2012. This resulted in the addition of 111 records of landslides events to the NLD over the two years investigated (2006 and 2012). We also find that we were able to obtain further information about landslide impact for 60–90 % of landslide events identified from newspaper articles. Spatial and temporal patterns of additional landslides identified from newspaper articles are broadly in line with those existing in the NLD, confirming that the NLD is a representative sample of landsliding in Great Britain. This method could now be applied to more time periods and/or other hazards to add richness to databases and thus improve our ability to forecast future events based on records of past events.

List of acronyms and abbreviations used in ***Chapter 3***

Abbreviation	Meaning	Section Introduced
BGS	British Geological Survey	3.1
LRIM	Landslide Road Impact Model	3.1
NHP	UK Cabinet Office Natural Hazards Partnership	3.1
NLD	National Landslide Database of Great Britain	3.1
UK	United Kingdom	3.1

List of variables used in ***Chapter 3***

Variable	Meaning	Section Introduced
D	Date precision of landslide record (number of days either side of date in which landslide could have occurred)	3.4
L	Time lag between landslide occurrence and reporting (variable units)	3.4
N_{LS}	Number of landslides	3.3
S	Spatial precision of a landslide record (estimated radius from true landslide location in metres)	3.4

3.1 Introduction

This **Chapter 3** presents results from a six-month part time internship undertaken with the British Geological Survey (BGS) and the United Kingdom (UK) Cabinet Office Natural Hazards Partnership (NHP) in 2013 and 2014. This **Chapter 3** is published as Taylor *et al.* (2015), with only minor differences in text (i.e., this introduction and **Section 3.5.4** discussing the implications of the findings of this chapter on the development of the landslide road impact model, LRIM). The research was motivated by the need to develop an impact model for hazards that are thought to frequently affect the UK, of which surface water flooding, severe weather and landslides were selected (Met Office, 2013). Some of the main goals of the hazard impact model are to:

- *Improve our ability to forecast hazards*, and their interactions.
- *Build a library of previous impacts* from hazards.
- *Combine the forecast of hazard with measures of exposure* (population, infrastructure and the environment) and the library of past impacts to forecast and visualise in real time the likely impact of an event.

All three of the above aims require robust databases of the incidence, characteristics and impacts of past hazard events in the United Kingdom. For landslides, this information comes from the *National Landslide Database* (NLD) of Great Britain, which is held and maintained by the BGS (Foster *et al.*, 2008).

The NLD is a database of primarily secular landslides (i.e., individual landslides) that have occurred since the last glaciation (Foster *et al.*, 2008). In **Chapter 2**, we discussed issues with inventory and database methods which generally result in some landslides not being catalogued. Although the NLD does not make any claims about being a 'complete' record of landsliding in Great Britain, we show in this **Chapter 3** that by searching a digital

archive of local newspapers, we can add more records and richer information about landslides to the NLD. As we will show, the method we develop does not 'complete' the database by any means, but is a useful and relatively easy way to supplement the evidence base upon which the UK landslide hazard impact model is built. As this method to supplement existing records of landslides in Great Britain draws consistently upon an independent dataset (a digital archive of local newspapers), comparing the results to the contents of the NLD can also provide a way to assess potential bias in the NLD and enhance overall confidence in its data. The method we present here could also be applied to enhance understanding of other natural hazards, such as surface water flooding, whose incidence and impacts are not systematically recorded in existing datasets, particularly when examining records pre-remote sensing (Moores and Rees 2011; Hurford *et al.* 2012).

The experience of working with the BGS and NHP and discussing how to record landslide impact (**Section 3.4.4**) also helped inform this author's understanding of the infrastructure disruption caused by landslides and how this can be measured. By these means, this **Chapter 3** can be considered part of the formative background in the development of the Landslide Road Impact Model (**Chapter 7**).

In the following sections of **Chapter 3** we:

3.2 *Summarise the broader difficulties of producing landslide inventories*. Discuss the potential of newspaper articles as a supplementary source of landslide inventory data. Introduce the particular newspaper archive used in our research.

3.3 *Describe the methodology* we developed for searching and filtering digital archives of regional newspaper to collect news stories about landslide events and extract factual information from them to enrich the NLD.

3.4 *Present results* of our newspaper searches for two search periods.

3.5 *Discuss the implications and uncertainties of our methodology* and how this methodology might be applied in other contexts.

3.6 *Summarise* results and draw conclusions.

3.2 Background

3.2.1 Landslide Inventories and Databases

Detailed information about the nature of past events is important for understanding, predicting and managing landslide risk (Guzzetti *et al.*, 2005; Guzzetti *et al.*, 2012). Van Westen *et al.* (2006) identify four basic types of information about past landsliding needed to support risk assessment and management:

(i) *Inventories* of landslides.

(ii) *The environment* surrounding the landslide.

(iii) *What triggered* the landslide.

(iv) *What elements* are/were at risk.

Of the four categories given above, van Westen *et al.* (2008) and Van Den Eeckhaut and Hervás (2012) demonstrate that the first category, landslide inventories, is the most important when considering potential risk for the future.

Compiling such inventories is complicated by a number of factors, discussed in detail in **Chapter 2** but including:

(i) *First order conceptual questions about the definition of a landslide 'event' to be recorded as distinct from a landslide triggering event* (e.g., an earthquake or heavy rainfall) (Kirschbaum *et al.,* 2010).

(ii) *Lack of direct measurements*. Compared to other hazards where we often have direct instrumental measurements of the phenomena over a wide region (e.g., ground motion, air temperature), landslide deposits observed on the ground are the *outcome* of a set of interacting processes (Guzzetti *et al.,* 1999) that are rarely feasible to measure systematically instrumentally. Consequently, to produce a landslide inventory, one must actively search for them across a landscape.

(iii) *Difficulty of identifying and extracting landslide events from public databases*. For example, in the UK the Highways Agency Road Impact Database, landslides do not have a specific event code. Landslides and engineered slope failures are sometimes noted in a free text field but are more commonly recorded in their database of traffic disruption as 'other' (Met Office, personal communication, March 2014).

For the above three reasons, it is rare to have databases of all landslides that have occurred over a region within a given time period, and there may be biases towards locations where humans are affected (Carrara *et al.,* 2003) or larger landslides that are more discernible in imagery/field studies (Willis and McCrink, 2002). The 'completeness' of an inventory will also be affected by the time lag between the landslides occurring and when they are inventoried, as smaller landslides may be eroded/erased from the landscape within a few months of occurring (Malamud *et al.,* 2004a, Bell *et al.,* 2012). In a survey of 22 European countries that have or are developing national landslide databases, Van Den Eeckhaut and Hervás (2012) found that 68 % of respondents estimated the completeness of their

country's database to be less than 50 %. The above difficulties with the completeness of landslide inventories limit the quality and predictive power of landslide susceptibility assessment (Galli *et al.,* 2008). Consequently, landslide risk may be under or overestimated depending on the completeness and homogeneity of coverage of the landslide inventory.

3.2.2 The National Landslide Database (NLD) of Great Britain

The National Landslide Database of Great Britain (NLD) is the most extensive source of information about British landslide occurrence. A metadata description with examples of its content can be found online at BGS (2014a). The NLD currently contains over 16,500 records of individual landslides occurring between the last glaciation and present day. For each landslide up to 35 possible attributes are recorded (Foster *et al.*, 2008; Pennington *et al.*, 2015). These can broadly be categorised into:

- *Landslide location* (Latitude/Longitude and estimation of locational precision).
- *Landslide timing* (date of occurrence or age).
- *Type of landslide* (fixed categories).
- *Cause of landslide* (fixed categories).
- *Size of landslide* (free text).
- *Impact of landslide* (number of fatalities, number injured, cost and other free text).
- *Geological setting of landslide* (fixed categories).

Perhaps due to the somewhat episodic nature of landslide activity in Great Britain, policy concern for landsliding has waxed and waned (Gibson *et al.,* 2013), as have resources for NLD data collection and database maintenance, resulting in temporal and spatial variations in database richness. The first national landslide database was initially established in the early 1980s to raise awareness of the nature and distribution of landslides for planning purposes at a local authority level (Foster *et al.,* 2012). As the method employed was a desk-based review of secondary sources such as technical reports, theses, maps and diaries (Jones and Lee, 1994), the spatial extent of records in the original NLD were biased towards locations of human interest, such as high impact landslides or 'classic' field study locations. During the 1990s, sources of revenue from the database were not large enough to fund the maintenance and regular updating of the database and the project was mothballed. In the early 2000s, the Department of Environment made the database available to the BGS, who over the next few years devoted considerable effort to restructuring, quality controlling, and supplementing this database into a more user-friendly and commercially relevant resource (Foster *et al.*, 2012). As of 2006, the NLD can be considered to be in its 'contemporary' phase, where information about new landslide

events are systematically recorded and added in 'live' (i.e., as and when the BGS hear about a landslide rather than through periodical retrospective studies). In addition to landslides occurring under natural conditions, since 2012 the BGS also records information about failures in engineered slopes, as they often cause considerable human impact (e.g., if a landslide occurs on a railway embankment, this could cause transportation disruption).

Information about landslides is added to the NLD through a number of primary and secondary research channels, which are described in detail in Foster *et al.* (2012) and Pennington *et al.* (2015). These can broadly be separated into:

- *BGS maps and archive documents* (e.g., field notebooks).
- *BGS field surveys*/reports.
- *Academic literature* (books, journal articles, student theses, etc.).
- *Aerial photography*.
- *Searches of archive media documents* (newspapers).
- *Online keyword searches of current media* sources (newspapers, radio, television, internet).
- *Personal communication* (public, local authorities, land owners, utilities operators).
- *Keyword searches of social media* (Facebook and Twitter) implemented since August 2012.
- *Citizen science reporting* via the BGS 'report a landslide' web-portal (BGS, 2014b) since 2009 and BGS Twitter profile (@BGSLandslides), implemented in 2012.

From 2008 to 2013, the search of current media which helps inform the NLD was performed by Meltwater (2014). Meltwater is a subscription media monitoring service aimed primarily at assisting organisations to manage their PR by scanning online media. They provided the BGS with a daily report based on the results returned from an automated Boolean search of a database of 190,000 online sources, including news, social media and blogs (Meltwater, 2014). However, the actual sources searched and how they may have changed over time are commercially confidential.

With the rise of social media, Twitter has become along with traditional media reports, a primary channel by which the BGS is alerted of landslide events. Where possible, alerts are followed up via field investigation or contact with affected groups/land owners, prior to inclusion in the NLD. Pennington *et al.* (2015) estimate that addition of social media and inclusion of engineered slope failures since 2012, and improved traditional media search strategies, have increased the number of NLD additions per year by a factor of 10 compared to the start of the contemporary phase (2006).

In the following **Section 3.2.3**, we describe the use of newspaper articles as a source of information about landslide events. In **Section 3.2.4**, we introduce the Nexis UK archive of regional newspaper stories and discuss differences between the current media search strategy used by the BGS and that of Nexis UK.

3.2.3 Newspaper Articles as a Source of Information about Hazards

Mass media is generally the first and primary source of information about hazards for the public (Fischer, 1994). Yet, mass media is also used by scientists and practitioners in the field of hazards in a number of ways, with varying levels of depth of engagement with the media:

(i) *First Alert.* A news article may be the first way a practitioner hears that a hazard event has happened. From this first alert, s/he may decide whether any follow up is required (such as a field visit) (e.g., GDACS, 2014; Pennington *et al.*, 2015; Public Health England, 2014).

(ii) *Archives (and Scientific Analysis of Archives).* Archives of news stories about various events can be searched to create or add to a database or inventory of hazard occurrence (e.g., Guzzetti *et al.*, 1994; Black and Law, 2004; Llasat *et al.*, 2009; Kirschbaum *et al.,* 2010).

(iii) *Documenting Impacts.* Media can be used as a way of documenting impacts of events from desk based studies, both at the time of occurrence and through future updates/press releases and reports (e.g., Tarhule, 2005 for droughts/floods, and Petley *et al.*, 2007 for landslides).

(iv) *Public Perception of Risk.* Analysis of the interactions between mass media coverage and public understanding of hazards and risk can be performed (Kasperson *et al.,* 1988). For example, media coverage of a particular hazard can be assessed over time to understand changes in how issues such as responsibility are framed (Escobar and Demeritt, 2014) or assessing variation in interest in a particular story over time (Carvalho, 2007).

(v) *Public communication.* Information can be disseminated through interviews and press statements (creation of media) (Peters *et al.,* 2008).

The use of newspaper articles as a proxy for records of various hazards is not a new technique. In a review of proxy records, Trimble (2008) lists examples of studies as early as 1932 using newspaper reports to construct a record of major landslides occurring in Switzerland from AD 1563 onwards (Heim, 1989 [1932]) and in 1946 using newspaper

reports to reconstruct a record of flooding in Utah (Woolley *et al.*, 1946). The technique is also well established in historic climate reconstruction (Demeritt, 1991; Brázdil *et al.*, 2005).

Raška *et al.* (2014) provide an overview of natural hazard databases that use newspaper and other documentary evidence. For landslides, perhaps the most cited national database is the Italian AVI project (available online at CNR–GNDCI, 2012), containing records of > 32,000 landslide and > 29,000 flood events, going back 1,000 years, but with most recorded between 1900 and early 2000, of which ~78 % of the information comes from newspaper reports (Guzzetti *et al.,* 1994; Guzzetti and Tonelli, 2004). More recently, the growing capacity to search freely available digital archives of global newspaper reports and online sources has prompted the construction of the Durham Fatal Landslide Database, which includes fatal rainfall triggered landslide events occurring across the globe since 2004. For the seven year period, 2004−2010, the database includes 2,620 fatal landslides, which resulted in 32,322 fatalities (Petley, 2012). Other examples of landslide databases using newspaper articles as a source of information include Domínguez Cuesta *et al.* (1999) in the North of Spain, Glade and Crozier (1996) in New Zealand, Devoli *et al.* (2007) in Nicaragua and Kirschbaum *et al.* (2010) at the global scale.

There are clear biases in newspaper articles as a proxy for information about hazards, such as an overemphasis on events with human impact (Carrara *et al.*, 2003), increased media interest following a number of events (Pennington and Harrison, 2013), a focus on high magnitude events or underreporting of low magnitude events (Guzzetti and Tonelli, 2004) and scientific correctness of information (Ibsen and Brunsden, 1996). Nonetheless, the regular publishing intervals (and thus continuous record) (Raška *et al.*, 2014) and relative ease and low associated costs of performing a desk-based study means that analysis of newspaper articles is widely seen as a useful compliment to other methods for building hazard databases. For example, in a review by Tschoegl *et al.* (2006) of 31 major international, regional, national and sub-national hazards databases, newspaper reports are used as a regular and/or major source of records about hazard events in 10 of the databases.

3.2.4 The Nexis UK Regional Newspaper Archive

In the last decade, there have been considerable advances in the digitisation and indexing of archives of newspapers in the UK, for example, The British Newspaper Archive (British Library, 2014) and The Nineteenth Century Serials Editions (NCSE, 2007). Here we have

been exploring the use of a digital subscription archive, Nexis UK (LexisNexis Academic, 2014), to add richness to the National Landslide Database of Great Britain (NLD). The archive was chosen due to its national scope, coverage up to present day and relative ease of searching. The method described in the following sections could be applied to other archives and extended back in time, as we will discuss in **Section 3.5**.

The Nexis UK archive of regional newspapers contains records of the print versions of 568 newspapers from across the United Kingdom (England, Wales, Scotland and Northern Ireland). For our purposes, we focus on the information that can be extracted from them to enrich the NLD which covers just Great Britain (England, Wales, and Scotland).

Whilst Nexis UK coverage is continuous from 1998 to present (LexisNexis Academic, 2014), some selected newspapers have records going back further, although Deacon (2007) cautions that there are some small inconsistencies in how data has been archived. For storage reasons, the Nexis UK archive does not include any original photographs from the news story, so some potentially useful information is lost (Weaver and Bimber, 2008).

Although national newspapers are also archived within Nexis UK, we decided to focus efforts on UK regional newspapers rather than national ones. By their nature, most landslides are local events with local impacts that would be newsworthy at a local to regional level. Any landslides large enough (or with extensive enough impacts) to make the national news would most likely also be captured in the local press.

At the time of undertaking this research, the British Geological Survey (BGS) had already used media sources (e.g., through Meltwater) to add information to the NLD. However, there are distinct differences between the media sources used by the BGS and the large archive of regional newspapers, Nexis UK, proposed here. Both sources are digital online services. Meltwater is a record of online news, whereas Nexis UK is a record of printed news. Even if both Meltwater and Nexis UK return records from the same newspaper, the content and length of the stories may vary. In an example given by Greer and Mensing (2006), a study comparing coverage of a news story about genetic cloning across three national broadcast news websites and three national newspaper websites, researchers found that online news stories were generally 20–70 % shorter, around 50 % of stories were written by newswire services (compared to 10 % in print) and generally the websites contained fewer citations. It is not clear how many of the regional newspapers included in Nexis UK database also have an online outlet that is being searched by Meltwater, but it is

clear that the content may well differ between the two, and as we will show, the Nexis UK database adds a large number of 'new' records of landslides to the NLD.

3.3 Methodology

In this **Section 3.3**, we present our methodology for searching the Nexis UK archive of regional newspapers to enhance the National Landslide Database (NLD). This process involves five major steps (**Figure 3.1**):

A. *Construct a set of Boolean search terms* to query the Nexis UK archive (outlined in sub-steps **A1–A5**)
B. *Apply the search terms* to obtain all articles from a given time period to return a corpus of potentially relevant articles
C. *Skim-read each article* from this corpus to identify those which are relevant
D. *Identify whether relevant article refers to a landslide already recorded within the NLD*
E. *Extract and code relevant information from the relevant articles*
F. *Pass information on to BGS for quality assurance, cross-checking and NLD data upload*

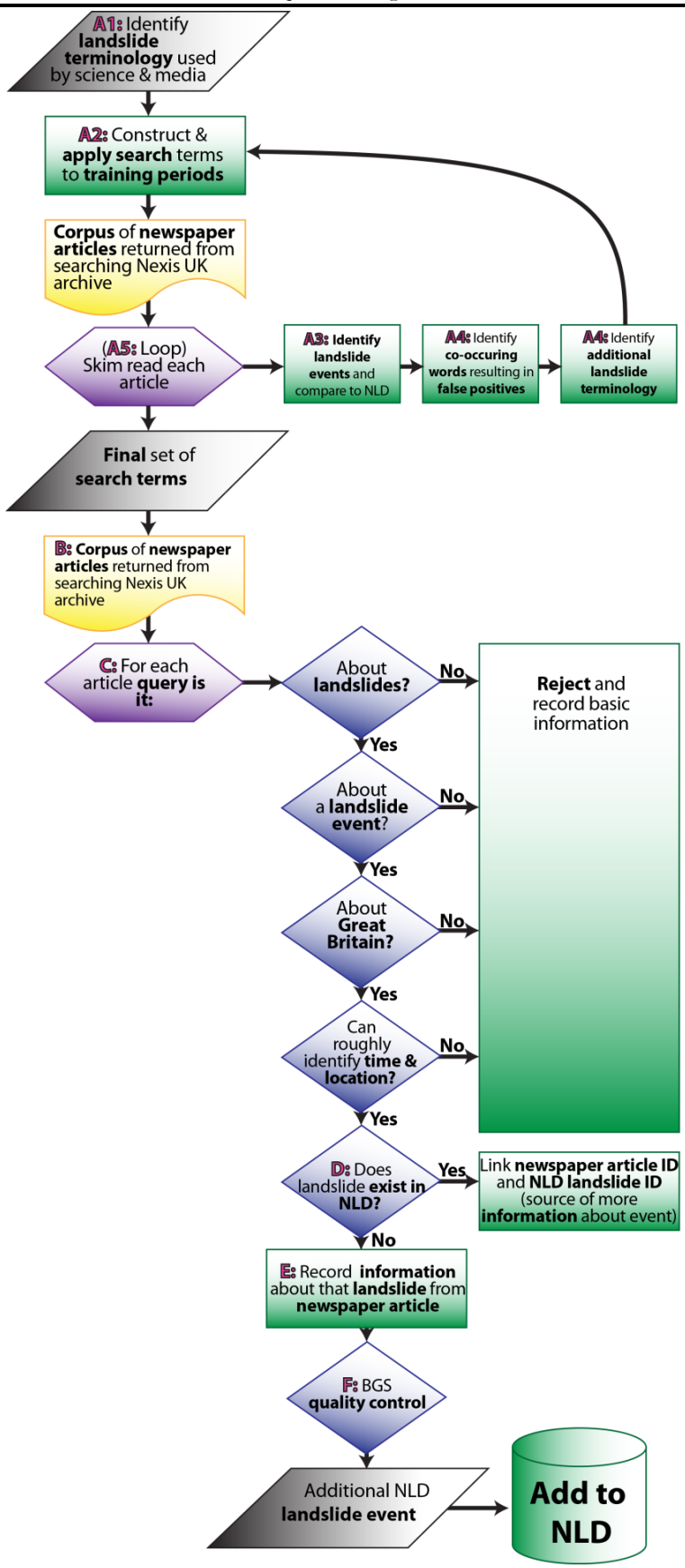


Figure 3.1 *Flowchart showing the principle stages of the methodology to create a systematic search of Nexis UK regional newspaper articles and extract information about landslide events to add to the richness of the British Geological Survey (BGS) National Landslide Database (NLD). The steps denoted in magenta letters and numbers (**A1** to **A5**, **B** to **F**) correspond to the steps identified in the text.*

3.3.1 Step A: Construct Search Terms

Nexis UK returns newspaper articles based on a Boolean keyword search (i.e., whether a word or combination of words does or does not appear within an article). There were multiple criteria for the search:

(i) *Maximise the number of articles about landslides in Great Britain*, particularly those that are lesser-known or unlikely to be recorded in the NLD.

(ii) *Minimise the number of false positives* (e.g., articles where the search terms appear in other irrelevant contexts such as 'a landslide victory').

(iii) *Ensure search terms capture any regional or temporal variation in landslide terminology*. For instance the Oxford English Dictionary (OED Online, 2014) notes that landslip is used chiefly in British English and thus would be a less appropriate search term for other parts of the world. Words also change over time, for example the word 'slough'. The OED notes one meaning for slough (pronounced slaʊ) comes from old English and connotes soft, muddy ground or mires, and another comes from Middle English (pronounced slʌf, sometimes also spelled sluff) and meant outer skin or peel. It was extended metaphorically by 19th century geologists to describe the surficial material shed by engineered embankments and steep scree slopes. In verb form the two meanings come together, insofar as the sloughing of rock or soil is usually *down* into a hole or depression.

Landslides are referenced using many different words by scientists, practitioners and the public, thus we use several Boolean search terms. To refine search terms satisfying the criteria listed above, five sub-steps were completed within **Step A**:

A1. *Identify key landslide terminology from the sciences and the media.*
A2. *Apply search based on A1 for selected training periods.*
A3. *Read through all resulting articles* from step A2. Identify landslide events and compare these to those already existing within the NLD.
A4. *Identify any additional terms* used to refer to landslides as well as co-occurring words associated with false positives.
A5. *Incrementally add the additional search terms* found in step A4 to the existing search terms in A2 and re-apply search. At each stage verify if any articles about landslide events are being filtered out and/or a large number of false positives are being added in.

Search Term Construction Stage A1 [Identify key landslide terminology]: Key landslide terminology outlined by Varnes (1978) and Cruden and Varnes (1996) were assessed, selecting the terms that are more commonly used in the English language or styles of landslide particularly

prevalent in Great Britain (**Table 3.1**, highlighted in **bold underline**). More commonly used terms known to be used in the British media were added, based on previous BGS experience of searching media, including 'landslip', 'slope failure' and 'slope instability'.

Table 3.1 *Landslide terminology for different styles of landslide in various materials from the Varnes (1978) and Cruden and Varnes (1996) classification system. Highlighted in* ***bold underline*** *are terms we deemed to be more commonly used in the English language and/or styles of landslide commonly seen in Great Britain.*

	Rock	*Debris*	*Soil*
Fall	**Rockfall, Rock fall**	Debris fall	Soil fall
Topple	**Rock topple**	Debris topple	Soil topple
Slide (rotational)	Slump	Single/Multiple/ Successive	Single/Multiple/ Successive
Slide (translational)	Block slide	Block slide	Slab slide
Planar	Rock slide	**Debris slide**	**Mudslide, mud slide**
Lateral Spreading	Rock spreading	Debris spread	Soil spreading
Flow	Rock flow	**Debris flow**	**Mudflow, mud flow**

Search Term Construction Stage A2 [Apply search based on A1 for selected training periods]: To test the robustness of the combination of search terms from **Stage A1**, they were applied to the Nexis UK archive of newspaper articles over four sample training periods: 1–31 December for 2004, 2005, 2006 and 2012. Landslide events during December 2012 had a high media profile, with events routinely recorded from national press and social media. During the years 2004 to 2006, 'live' data collection and recording of events were not so systematic, in addition to engineered slope failures and smaller events being rejected. These particular time periods were therefore chosen in order to test semantic variability over the range of the Nexis UK archive and to compare with and add richness to the NLD.

Search Term Construction Stage A3 [Read through all resulting articles from step A2 to identify landslide events]: Each newspaper article was skim-read to check whether it satisfied the following criteria:

- *Is the article relevant* (i.e., related to the geomorphic process of landslides)?
- *Is the article about a landslide 'event'* (rather than general discussion of landslides)?
- Is the article about a *landslide event that occurred in Great Britain*?
- *Is it possible to roughly locate and date the landslide* event (possibly requiring further desk-based research)?

If any of the four criteria above were not satisfied, the article was rejected and basic information about the article systematically recorded (to allow future database interrogation). If all of the four criteria above were satisfied, a search of all landslides already existing in the NLD was performed to check whether the landslide was already recorded. If the landslide was already recorded, the newspaper article ID was linked to the NLD landslide ID as a potential source of more information and confirmation. If the landslide event was not in the NLD, as much information as possible about the landslide was extracted from the article and systematically recorded using the same structure as the existing NLD (described in **Section 3.2.2**).

Search Term Construction Stage A4 [Identify any additional search terms]: All articles referring to landslides were read carefully to identify any additional terms for landslides used within the texts. This resulted in the additions 'cliff collapse' and 'land movement'. Variations of 'cliff collapse' were also added in ('coastline collapse' and 'cliff fall'). We also identified co-occurring words associated with false positives (i.e., articles about electoral rather than geologic landslides); all irrelevant articles were coded into themes, and key words selected based on these themes to modify the Boolean filter to remove any articles containing the words 'elect' (or derivatives such as elected), 'victory', 'win', 'won', 'majority', 'submarine' and 'porn'.

Search Term Construction Stage A5 [Incrementally add the additional search terms]: At each stage, the search of Nexis UK for the training periods was re-applied, and the resulting articles checked to verify that (i) no landslides previously identified were now being filtered out and (ii) no large number of false positives were being added in.

In this **Step A**, constructing the final set of search terms used in the rest of our research, there were two cases where a large number of irrelevant articles were returned. The decision was made not to filter results because this would inevitably filter out relevant articles. The first of these was 'cliff falls' – which captured reports about people falling from the top of cliffs as well as ones about the coastal cliff instability. Given the semantic overlap between these two reporting themes, automated methods could not distinguish between them easily, so it was decided to use manual ones instead. The second case included articles about landslide events occurring abroad (e.g., following a typhoon or earthquake in Asia). Nexis UK offers some additional search filters, such as searching by geography (articles tagged as referring to a specific country) and newspaper section (e.g., only returning articles in the 'News' or 'Music' sections). However, we chose not to use

these filters as sample testing showed that regional newspaper articles are not consistently classified in Nexis UK, therefore the results were too limiting. Manual filtering was used to deal with articles from regional newspapers in Northern Ireland, so as to only choose stories that referred to landslides in Great Britain.

The final search terms that we used for *all* subsequent searches are given below. This includes the use of Boolean logic (OR, AND, NOT) and wild cards (*, !) to search for different derivatives of given terms (e.g., Landslide* returns the words Landslide, Landslided, Landslides):

> [**Landslide*** OR **landslip*** OR **slope failure** OR **rock fall** OR **rockfall** OR **mudflow** OR **mud flow** OR **cliff fall** OR **slope failure** OR **slope instability** OR **debris flow** OR **land movement** OR **cliff collapse** OR **mudslide** OR **mud slide** OR **coastline collapse** OR **rock topple** OR **debris slide** AND NOT (*elect!* OR *victory* OR *win* OR *won* OR *majority* OR *submarine* OR *vote* OR *porn*)]

Terms in **bold underline** indicate that if one instance of that term appears, then the article will be flagged as a potential landslide relevant article. Terms in *italics* indicate that if an article contains any of the bold-underlined black words but also contains one of the italicised words, the article will be filtered out of search results. * = wildcard of 1 character; ! = wildcard of 1 or more characters.

Figure 3.2 shows results from applying the final search terms (**Step A5, Section 3.3.1**) to the four training periods (Decembers 2004, 2005, 2006 and 2012). For the December 2004 and 2005 test periods, the NLD did not have any records of landslide events, whereas four landslides were identified in each month using the Nexis UK archive. This demonstrates the potential value of applying the method outlined here to enrich the NLD for the period prior to 2006 period when the BGS entered its 'contemporary' phase of data collection. For the December 2006 test period, the NLD contains records of seven landslide events, three of which were also identified in Nexis UK articles. In this month, we also detected four additional landslide events not previously recorded in the NLD, representing a 57 % increase in database entries for December 2006 by using the Nexis UK archive as an additional source of information. December 2012 was part of a particularly wet season, resulting in many more reported landslides than usual (Pennington and Harrison, 2013). At the time of performing this research, there were 75 landslides in the NLD for December 2012. Of these, 24 % were also identified in the Nexis UK archive. We also detected an additional six landslides not recorded in the NLD, increasing the total number of landslide

events recorded for December 2012 in the NLD by 8 %. The decline between 2006 and 2012 in the proportion of landslides detected using the Nexis method but not currently existing in the NLD, can be explained by the addition of social media as a source of information and the subsequent inclusion of engineered slope failures in the database.

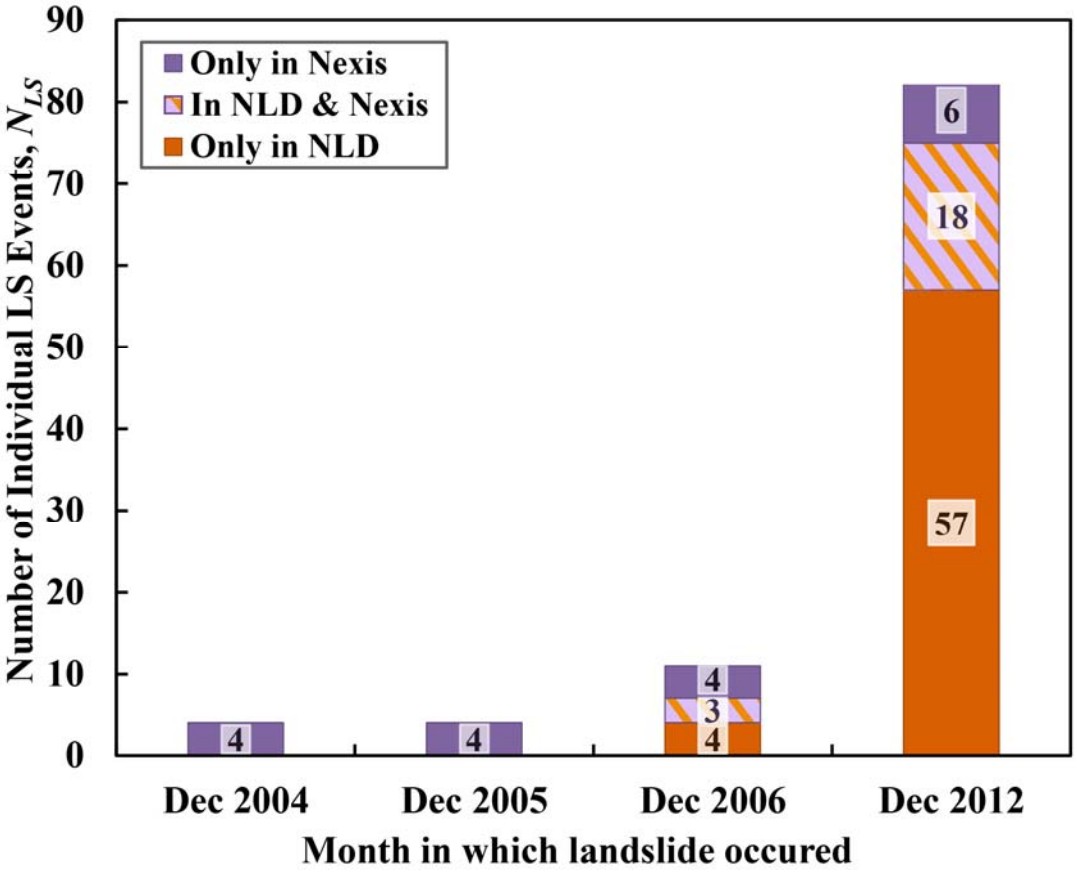


Figure 3.2 *Number of landslide events (N_{LS}) found from the Nexis UK archive search for the December 2004, 2005, 2006 and 2012 training periods that were: (i) only found in Nexis UK (i.e., they did not already exist in the NLD) (solid purple), (ii) found in both the NLD and Nexis UK (hashed purple and orange); (iii) found in the National Landslide Database (NLD) but not found in the Nexis UK archive (orange).*

In December 2012, there appear to be proportionally more events (57/81, i.e. 70 %) in the NLD that were not found in Nexis UK than in December 2006 (4/11, i.e. 36 %). This contrast was investigated for the December 2012 test period by examining the source of information for each landslide event that was found in the NLD but not in the Nexis UK newspaper archive. **Figure 3.3** shows a breakdown of these sources. The principal reason for these landslide events being in the NLD but not Nexis UK was that they were reported in the media after 31 December 2012. There is good reason to expect that many of these December 2012 events would have been detected using the Nexis UK archive, if instead of searching for a single test month, the time horizon for searching had been extended to overcome this lag time between an event occurring and a story being published about it. The second most frequent reason that we found for landslides not being identified in Nexis

UK is the source being an online newspaper article from the Newsquest Media Group. This group publishes some 300 local/regional newspapers, but only the print version of many of these newspaper titles is available to search in the Nexis UK archive. From our experience, the content and frequency of publishing vary considerably between the online and print versions. For instance, online news articles may be uploaded daily, whereas the paper is printed once per week, and neither the online nor print version contain all stories of the other, leading to discrepancies in the search results we generated using the Nexis UK method and the media scans provided to the BGS for the NLD by the Meltwater method.

There were a small number of cases where the source was available in the Nexis UK archive, but the specific article was not. This was confirmed by performing additional searches of Nexis using the title of the article and just searching a specific source. This can sometimes happen with freelance or newswire stories where the newspaper does not own copyright and cannot make it available for searching in Nexis UK (LexisNexis Academic, 2014). The majority of the remaining landslide events not identified in the Nexis UK archive search were from sources not available to search in the Nexis archive (e.g., social media, websites). No landslide events recorded in the NLD but not returned from the Nexis UK archive appeared to be caused by filtering/errors with the search terms. Although it is not possible to validate these results against the 'true' number of landslides that actually occurred in Great Britain in this period, it does appear that the search terms and method used here has relatively good agreement with existing records in the NLD and is also able to add richness by identifying additional landslide events.

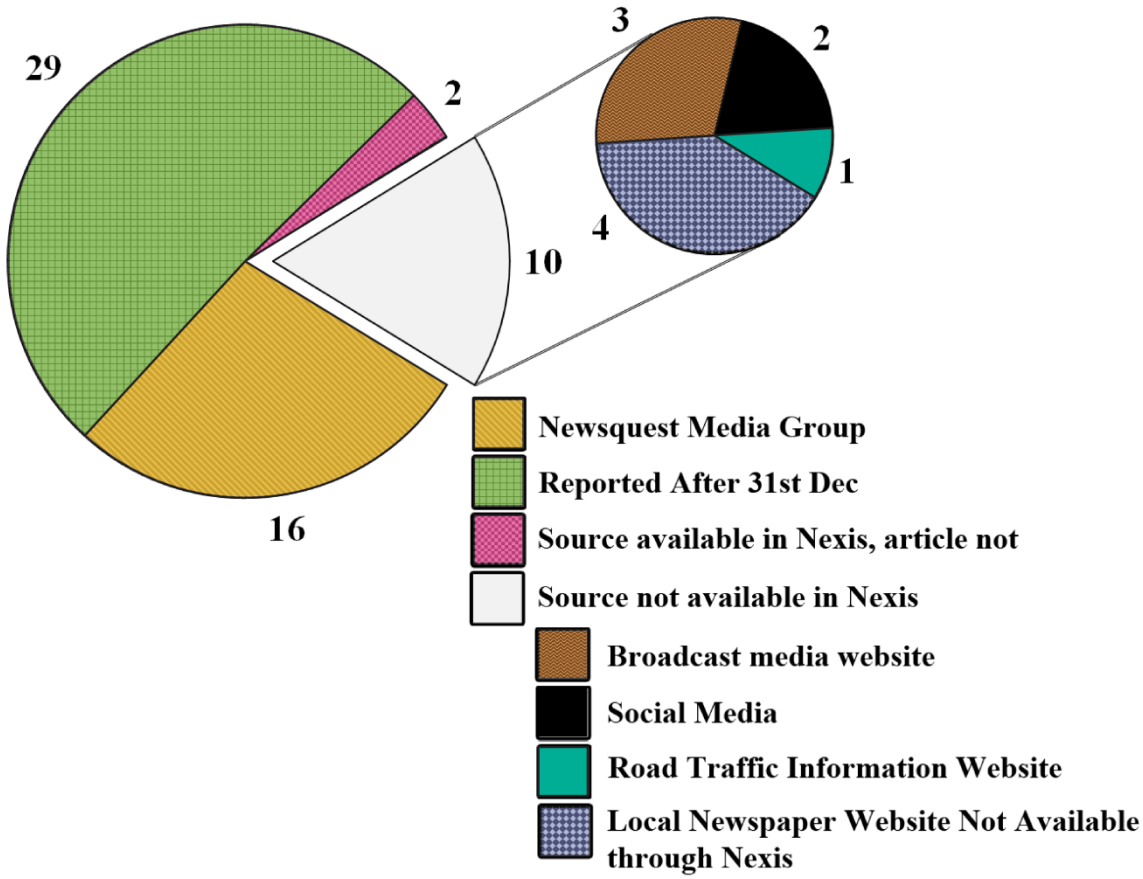


Figure 3.3 *Breakdown of sources for 57 landslide events noted in the NLD but not in Nexis UK for the December 2012 training period. Also included are 29 landslide events reported after 31 December 2012 (green).*

We did not identify any particular regional or temporal variations in landslide terminology. However, all test periods are relatively recent. It is possible that if the search was applied to more historical archives that spatial or temporal trends may appear in the landslide terminology used.

3.3.2 Step B: Apply Search Terms

The search terms (**Step A5, Section 3.3.1)** were applied to two time periods in the Nexis UK archive: all articles published between 1 January and 31 December for both 2006 and 2012. Once the search was applied, all newspaper articles were downloaded and input into a database to aid categorisation, creating a corpus of potentially relevant stories (see **Table 3.2** for the metadata recorded from two newspaper examples).

Table 3.2 *Examples of metadata for two newspaper articles returned from searching the Nexis UK archive of regional newspapers for articles published in December 2012. Articles are categorised depending on their relevance and whether they already exist in the NLD or not. (N = No, Y = Yes)*

Month, Year	Article ID	Title	Relevant?	Article Type	Enough Info to Locate?	Landslide in NLD?	Event ID in NLD	New Event ID	Full Text
Dec-12	1	Don't stop? "Fleetwood Mac will tour until we drop dead" says Stevie Nicks	N	I (Irrelevant)	-	-	-	-	"Do not call it a comeback and don't even think of it as a farewell tour. After more than four decades making music and a 2010 tour, Fleetwood Mac will hit the …"
Dec-12	2	Great Christmas getaway hit as signal failure causes widespread disruption	Y	LE (Landslide Event)	Y	N	-	N_2012_DEC1	"...the express service made additional calls. In Lancashire, trains between Liverpool and Manchester were diverted because of a landslip near Warrington. Many …"

3.3.3 Step C: Skim-Read Results

The title of each article was skim-read to ascertain whether it was relevant. This is demonstrated in **Table 3.2** where article 1 on Fleetwood Mac is clearly irrelevant from the title and is thus rejected and categorised as 'I' (Irrelevant). If the title suggests the article could be relevant, the full text was read to locate and date the possible landslide. In some cases, further desk-based research was required to ascertain whether the article truly referred to a landslide event or not. For example, one newspaper article referred to a landslide but then further described the event as "a building collapsed into a construction site". In such examples, desk-based research was undertaken to identify the exact location of the event using tools including Google Earth time lapse imagery, Google Street View,

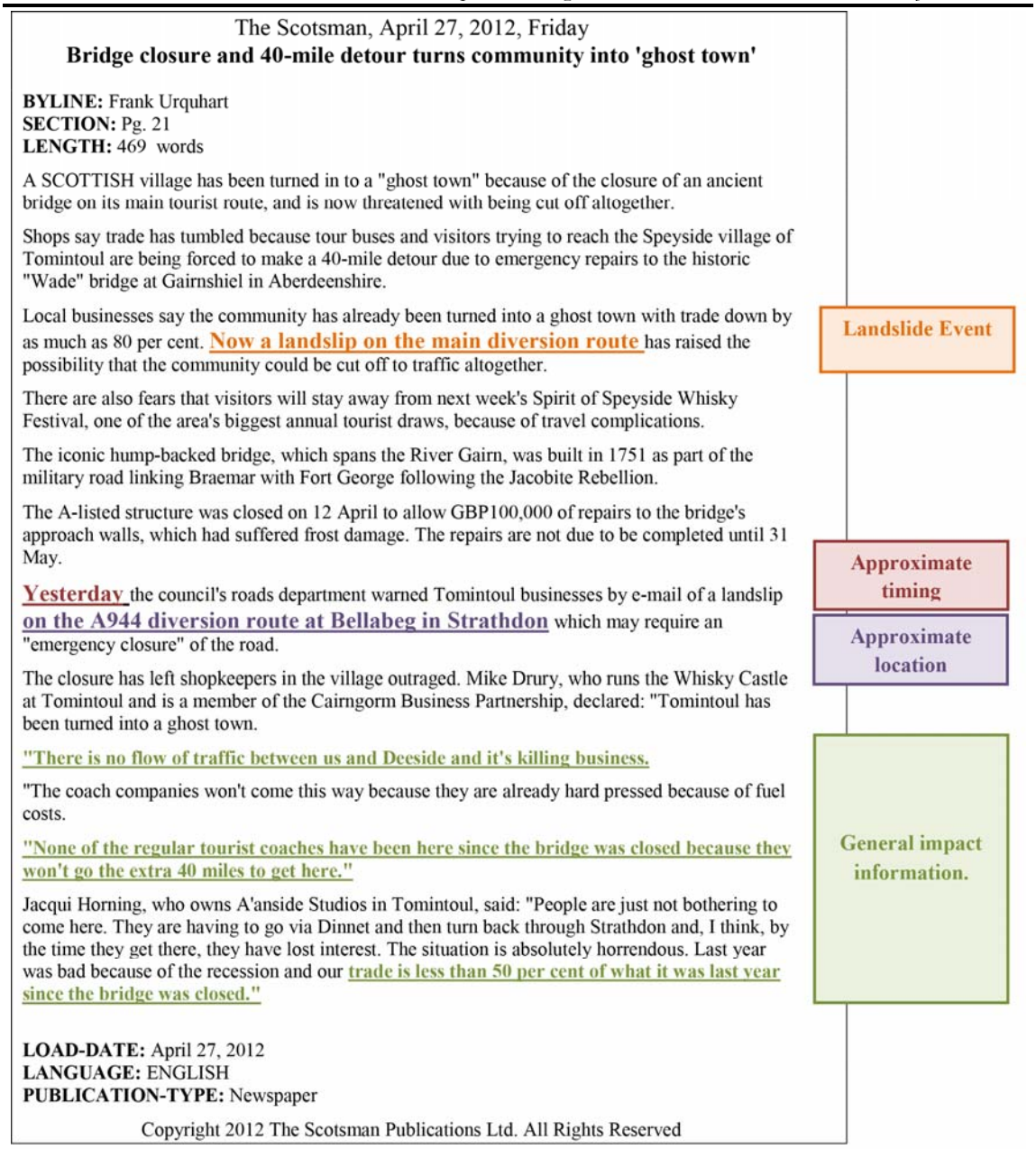
property websites, social media and other online sources to identify whether this was a landslide, a sinkhole, an issue with slope excavation or another type of event.

3.3.4 Step D: Identify Whether Relevant Article Refers to Event(s) in the NLD

As detailed in **Step A3**, if a relevant article contained enough information to approximately locate and date, a search was performed upon the existing NLD to see whether a record of the landslide existed. If so, the article was linked by ID to that landslide event, creating additional confirmation of this event and a potential source of further information to be processed at a later date. Newspaper articles containing more precise information (e.g., improved spatial precision), were used to update the original landslide event.

3.3.5 Step E: Extract and Code Relevant Information from the Article

If the landslide did not exist in the NLD, as much information as possible was extracted from the article and categorised according to the BGS NLD pro-forma and a case-by-case judgement of the precision of that information made. An example article is shown in **Figure 3.4**.



The Scotsman, April 27, 2012, Friday

Bridge closure and 40-mile detour turns community into 'ghost town'

BYLINE: Frank Urquhart
SECTION: Pg. 21
LENGTH: 469 words

A SCOTTISH village has been turned in to a "ghost town" because of the closure of an ancient bridge on its main tourist route, and is now threatened with being cut off altogether.

Shops say trade has tumbled because tour buses and visitors trying to reach the Speyside village of Tomintoul are being forced to make a 40-mile detour due to emergency repairs to the historic "Wade" bridge at Gairnshiel in Aberdeenshire.

Local businesses say the community has already been turned into a ghost town with trade down by as much as 80 per cent. **Now a landslip on the main diversion route** has raised the possibility that the community could be cut off to traffic altogether.

There are also fears that visitors will stay away from next week's Spirit of Speyside Whisky Festival, one of the area's biggest annual tourist draws, because of travel complications.

The iconic hump-backed bridge, which spans the River Gairn, was built in 1751 as part of the military road linking Braemar with Fort George following the Jacobite Rebellion.

The A-listed structure was closed on 12 April to allow GBP100,000 of repairs to the bridge's approach walls, which had suffered frost damage. The repairs are not due to be completed until 31 May.

Yesterday the council's roads department warned Tomintoul businesses by e-mail of a landslip **on the A944 diversion route at Bellabeg in Strathdon** which may require an "emergency closure" of the road.

The closure has left shopkeepers in the village outraged. Mike Drury, who runs the Whisky Castle at Tomintoul and is a member of the Cairngorm Business Partnership, declared: "Tomintoul has been turned into a ghost town.

"There is no flow of traffic between us and Deeside and it's killing business.

"The coach companies won't come this way because they are already hard pressed because of fuel costs.

"None of the regular tourist coaches have been here since the bridge was closed because they won't go the extra 40 miles to get here."

Jacqui Horning, who owns A'anside Studios in Tomintoul, said: "People are just not bothering to come here. They are having to go via Dinnet and then turn back through Strathdon and, I think, by the time they get there, they have lost interest. The situation is absolutely horrendous. Last year was bad because of the recession and our **trade is less than 50 per cent of what it was last year since the bridge was closed."**

LOAD-DATE: April 27, 2012
LANGUAGE: ENGLISH
PUBLICATION-TYPE: Newspaper

Figure 3.4 *Example of a newspaper article returned by searching the Nexis UK archive of regional newspapers for articles published in April 2012 and using the search terms given in* ***Step A5, Section 3.3.1****. This article refers to a landslide event and gives approximate information about the date, location and impact of that landslide.*

3.4 Results

In this **Section 3.4**, we present the results of applying the Nexis UK search method to all regional newspaper articles contained in the database published during the calendar years of both 2006 and 2012. In **Section 3.4.1**, we present the overall results of the search before detailed analysis of individual landslide events is undertaken. In **Section 3.4.2** we then describe how this method adds richness to the National Landslide Database (NLD) through finding previously undetected events and the addition of information to existing events (**Section 3.4.3**). Finally, in this section, we discuss the precision to which this

information can be estimated from newspaper articles (**Section 3.4.4**). In **Section 3.5**, we will discuss the reliability of this information and potential further applications of the method.

3.4.1 Search Results by Type of Article

The Nexis UK regional newspaper archive was searched using the terms listed in **Step A5, Section 3.3.1** for all articles published between 1 January and 31 December during 2006 and 2012. The initial search (**Step B**, **Section 3.3.2**) resulted in 711 articles in 2006 and 1,668 articles in 2012. All articles were then skim read and categorised into broad types (**Step C**, **Section 3.3.3**), which are listed in **Figure 3.5**. For both periods, around 20 % of articles were categorised as completely irrelevant (i.e., false positives), and around 20 % of articles were categorised as 'general landslide discussion', meaning they referred to landslide phenomena but were not specifically about any particular landslide event.

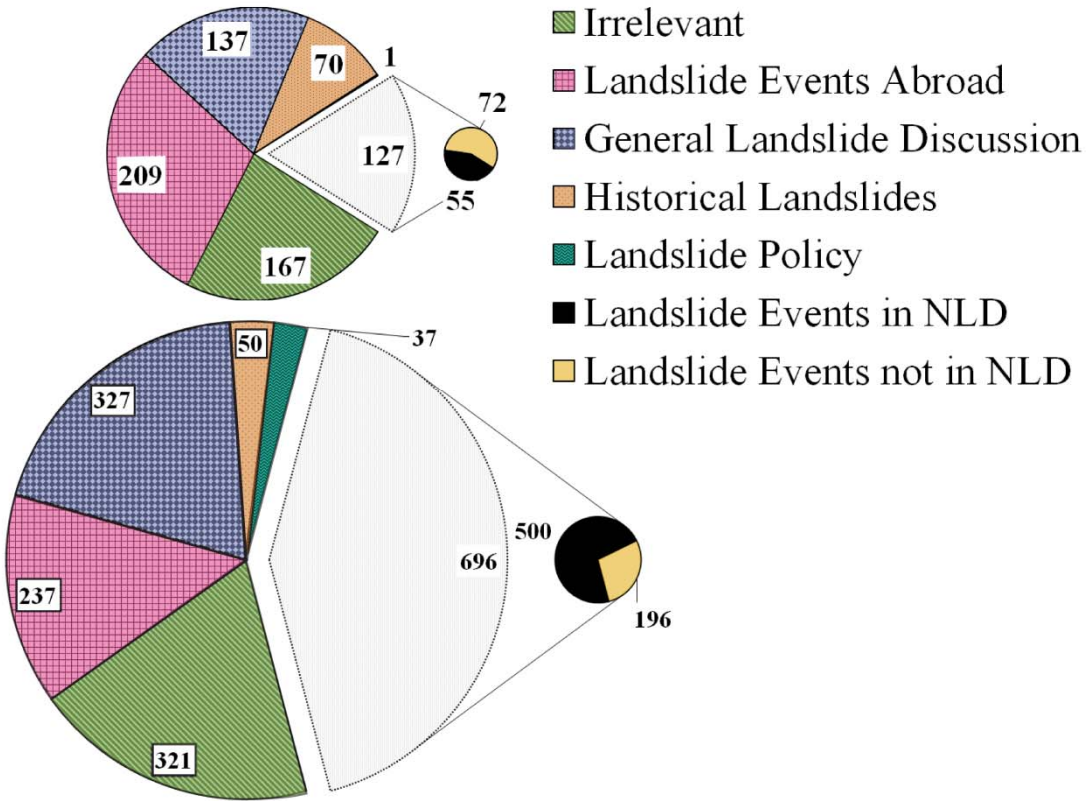


Figure 3.5 *Number of newspaper articles returned by searching the Nexis UK archive of regional newspapers broken down by type (e.g., articles that are irrelevant, or those that contain relevant information about a landslide event).* ***(A)*** *Results from 711 articles returned from the search of articles published between 1 January 2006 – 31 December 2006* ***(B)*** *Results from 1,668 returned from the search of articles published between 1 January 2012 – 31 December 2012.*

Broadly, there was a decline in the number of articles discussing landslide events abroad (outside of Great Britain) and historical landslides (those occurring before 2006) between

2006 and 2012. This is countered by an increase in the proportion of 'relevant' articles referring to a landslide event occurring in Great Britain, which rose from 18 % in 2006 to 42 % in 2012. This is possibly due to the fact that 2012 was a record year for landslides in Great Britain, resulting in increasing public and media interest (Pennington and Harrison, 2013). There was also an increase in the number of articles discussing landslide related policy in 2012. This is largely attributable to relatively unusual high-impact events occurring in 2012, such as fatalities, region-wide railway delays and repeated closure of stretches of road such as the A83 road at Rest and Be Thankful (Scotland), resulting in questioning from the press about what should be done to prevent landslides from a policy perspective. A similar effect has been noted in post-flood event coverage (Escobar and Demeritt, 2014).

Relevant articles referring to landslide events in Great Britain were then analysed more closely to associate them with particular landslides and extract information about those events with which to enrich the NLD in two ways:

(i) *Adding landslide events not previously recorded* in the NLD.
(ii) *Capturing more information* about landslide events already in the NLD.

In the following section, we discuss these two ways of enriching the NLD, starting with (i) additional events and their spatial patterning before turning to (ii) the additional information that our method of searching Nexis UK can generate about events already recorded in the NLD.

3.4.2 Adding Landslides to the NLD

Although we found 268 news articles referring to landslides not previously recorded in the NLD, many of these articles were referring to the same, rather smaller, subset of events. Once this repetition in our corpus of articles was accounted for, the final number of additions to the NLD was 39 events for 2006 and 73 events for 2012. This represents a 122 % and 40 % increase in the number of landslide events in the NLD for 2006 and 2012 respectively. We attribute these NLD additions principally to more and different sources now being searched, along with the majority of new landslides being relatively small in size and thus only of interest to the community in the immediate vicinity. **Figure 3.6** shows the number of additional landslide events per month for both years. In both years, the temporal trend in number of landslides per month is roughly the same: high landslide occurrence in the winter, and also a peak in the mid-summer. The pattern in number of additions from the Nexis UK method appears to vary between the years. In 2006, the

percentage increase in number of landslides added to the NLD per month varies between 33 % and 600 % and there does not appear to be a strong relationship between the number of landslides already in the NLD and the number of additions. Whereas in 2012, the percentage increase in number of landslides per month varies less (11 % – 300 %) and appears to be more strongly linked to the number of landslides already in the NLD for a given month. This suggests that the existing NLD is a representative sample of the temporal patterns of landsliding in Great Britain and that the BGS's development of search methods has been effective. Moreover, these results suggest that there is no strong bias for the month landslides are reported in by the media (e.g., in the summer months when there is relatively little political news), although testing of more years of data would be required to confirm this.

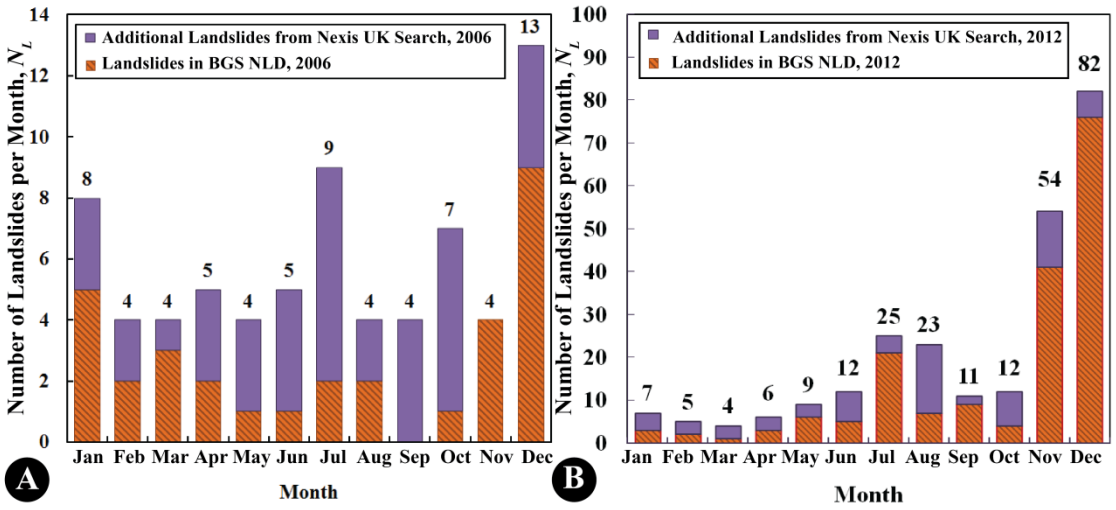


Figure 3.6 *Number of landslide events per month (N_{LS}) separated into (i) those already existing in the NLD (Orange hashed bars) and (ii) additional landslide events identified by searching the Nexis UK archive of regional newspapers (purple solid bars). Shown are results for 1 January to 31 December for:* ***(A)*** *2006 (BGS NLD: 32 landslides; additional landslides from Nexis: 39 landslides) and* ***(B)*** *2012 (BGS NLD: 178 landslides; additional landslides from Nexis: 72 landslides.*

Figure 3.7 shows separately for 2006 and 2012, the spatial distribution of landslide events already recorded in the NLD at the time of this research, and additional landslides added based on Nexis UK news coverage. The pattern in both years is broadly similar, suggesting no shift over time in the detection biases of this method. The distribution of events previously recorded in the NLD roughly matches that of the additional events detected from the Nexis UK regional newspaper archive but not yet recorded in the NLD. In both 2006 and 2012, landslides are strongly clustered in the South West of England, with smaller clusters in the North West (Yorkshire Dales), North Wales and the Highlands of Scotland; these areas of significant activity can be directly related to rainfall patterns and topography.

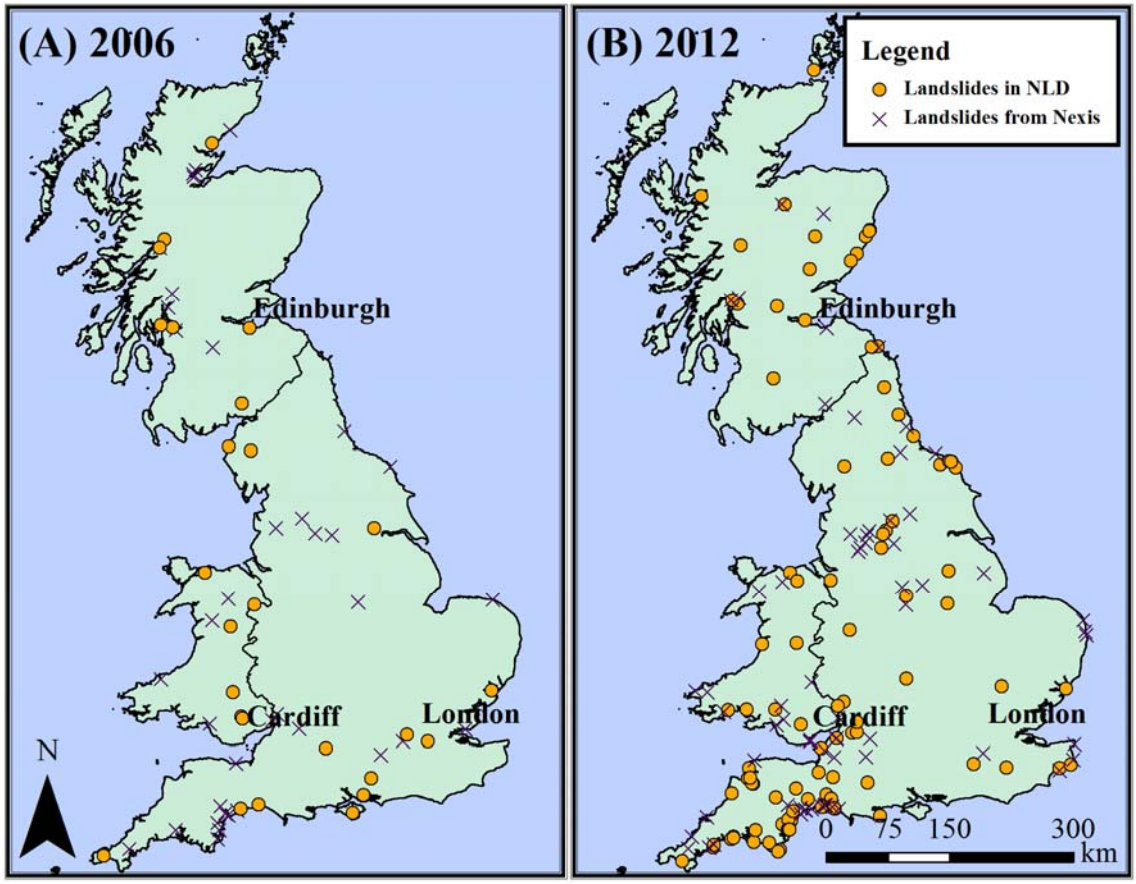


Figure 3.7 *Spatial distribution of landslides occurring 1 January to 31 December in* ***(A)*** *2006 and* ***(B)*** *2012. For both* ***(A)*** *and* ***(B)****, landslides are separated into (i) all landslides that occurred in 2006 and 2012 that already existed in the NLD at the time of this research (open circles) and (ii) additional landslides identified by searching the Nexis UK archive of regional newspapers (cross symbols).*

In **Figure 3.8** we show the spatial distribution of the combined landslides from 2006 and 2012, again for both landslides in the NLD at the time of this research, and additional landslides from Nexis UK, overlaid on a map of landslide susceptibility created from records within the NLD (BGS, 2014c). Broadly, the spatial extent of additional landslides correlates with regions of medium to high susceptibility in the existing susceptibility map.

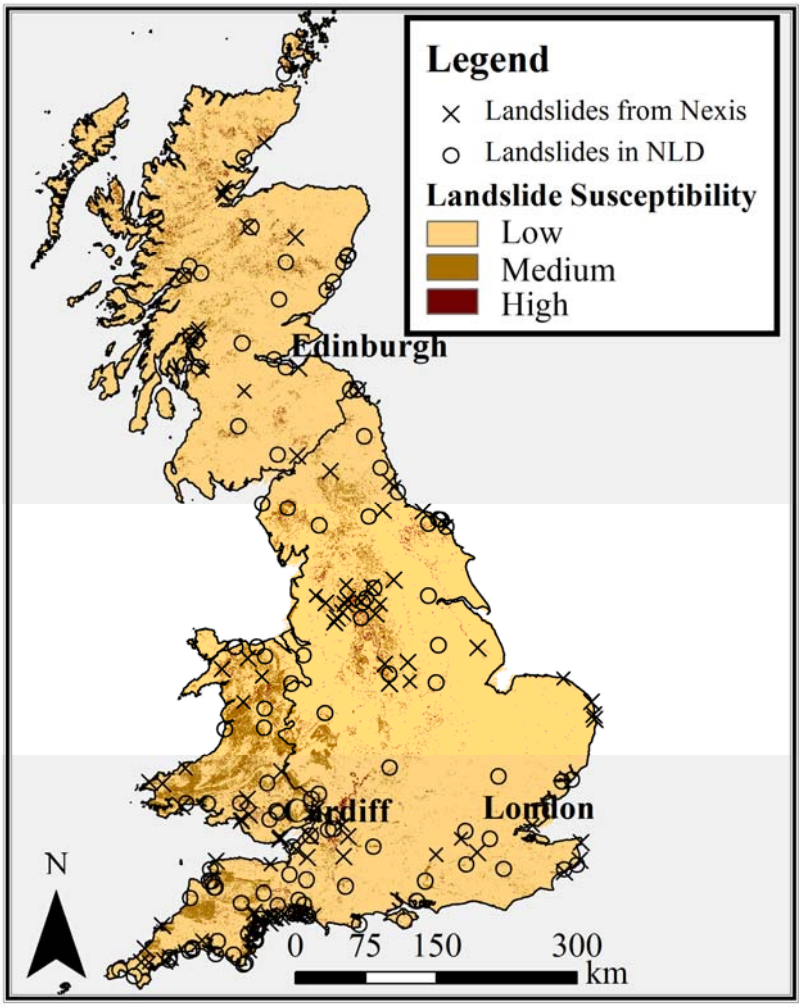


Figure 3.8 *The spatial distribution of landslides that occurred 1 January to 31 December in both 2006 and 2012 overlaid with a map of landslide susceptibility produced at 1:50,000 scale (BGS, 2014c). Landslides are separated into (i) all landslides that occurred in 2006 and 2012 that already existed in the NLD at the time of this research (open circles) and (ii) additional landslides identified by searching the Nexis UK archive (crosses).*

3.4.3 Capturing More Information about Landslides

As well as adding new landslide events to the NLD, the corpus of relevant stories generated by searching Nexis UK was also mined to enrich the NLD by capturing additional information about landslide events. As noted in **Section 3.2.1**, the existing BGS pro-forma records > 35 attributes (Foster *et al.,* 2008).

For ten landslide events (five in 2006 and five in 2012), additions and amendments were made to the records already in the NLD based on information included in Nexis UK articles. This included more precise dates and locations and additional impact information. Moreover, there are now 55 and 500 additional newspaper articles for 2006 and 2012 respectively that are linked to individual landslide events by ID, acting as additional confirmation for that event and a potential source of further information to be mined at a later date.

Figure 3.9 shows a breakdown of the type and/or availability of information available from newspaper articles for each additional landslide event identified from the Nexis UK search (n = 111), compared to the types of information available from a subset of the NLD (from 2006–2013, n = 471 at the time of doing this research). Newspaper articles are a good source of information for landslide date, approximate location and description of impacts. However, newspaper articles rarely contain more 'geotechnical' information such as the type of landslide, trigger and size. Elliott and Kirschbaum (2007) highlight the difficulty in classifying the type of landslide. Generally, landslide type classification was only possible from the articles in the Nexis UK archive for rock falls, which can be attributed to the relative simplicity of descriptions of large boulders rolling/detaching versus the more visually subtle difference between a planar/rotational slide.

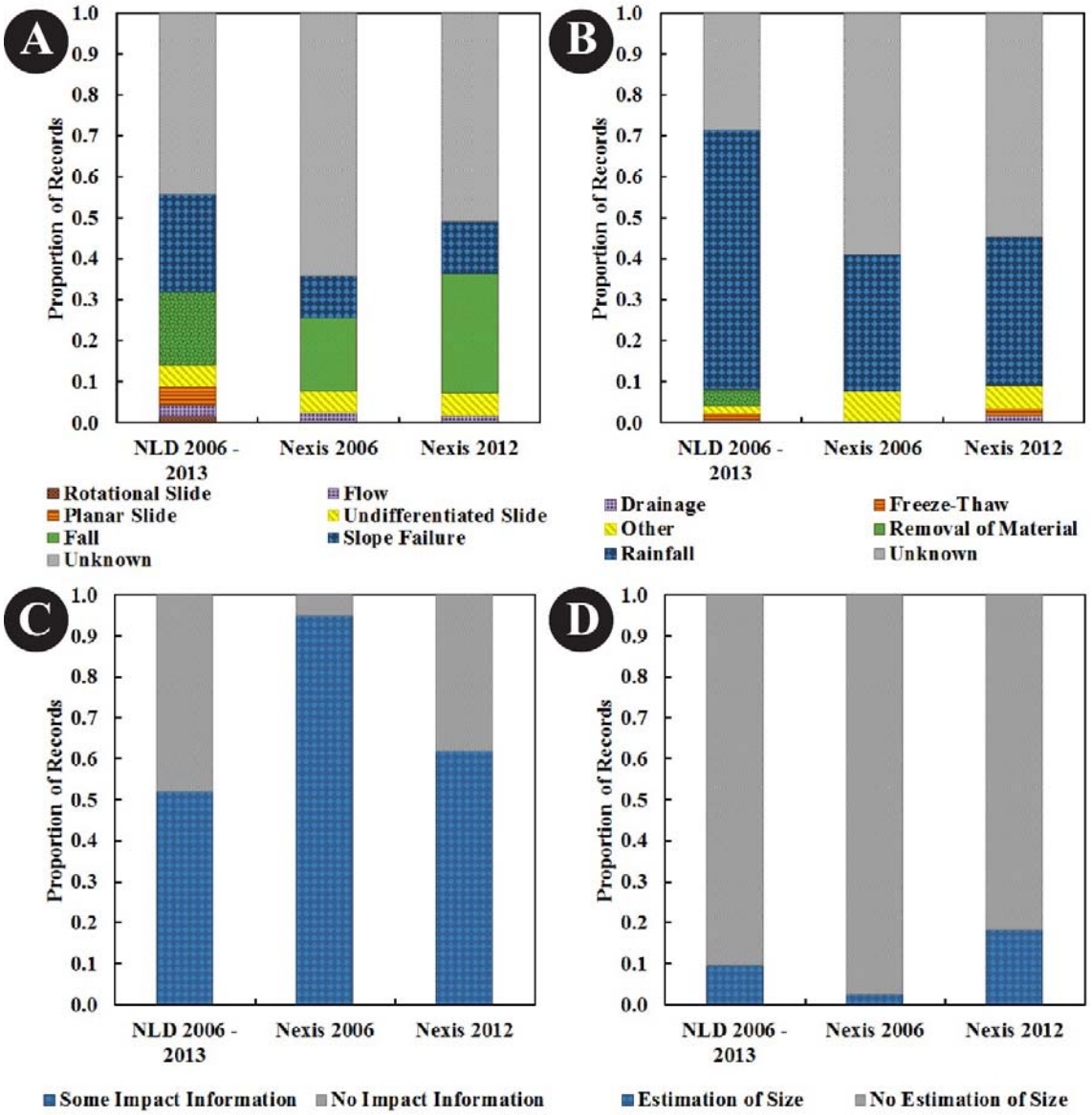


Figure 3.9 *Type and/or availability of information for each additional landslide event from the Nexis UK archive for 1 January to 31 December of 2006 and 2012 (n = 111), compared to the type and/or availability of information available for subset of landslide events already existing in the NLD for the eight year period, 2006 to 2013 (n = 471 the time of this research).* ***(A)*** *Type of landslide using fixed categories based on Varnes (1978) and Cruden and Varnes (1996) classification.* ***(B)*** *Trigger of landslide (fixed categories defined by the BGS).* ***(C)*** *Whether any information is available about the impact of the landslide (e.g., fatalities, injured, or other free text).* ***(D)*** *Whether any indication of the landslide size is given (predominantly free text).* ***Figures 3.9C*** *and* ***3.9D*** *are presented as binary categories due to the predominantly free-text nature of these data fields.*

Figure 3.9B shows that a trigger for a given landslide event could be identified from newspapers in less than half of cases. Typically the only trigger that could be inferred from an article was heavy or prolonged antecedent rainfall, which articles often described. Our findings based on newspaper articles are broadly consistent with the NLD, which indicates that 63 % of landslides in Great Britain were triggered by rainfall. It seems likely that many of the landslides from the Nexis search method missing this information were quite possibly triggered by rainfall.

Newspapers could also be mined for information about the impacts and size of landslides. As these are primarily 'free text' rather than categorical fields in the NLD, results are presented in binary terms of whether information was present or not. **Figure 3.9C** highlights the relative success of extracting landslide impact information from newspaper articles. As mentioned previously, this is most likely due to preferential coverage of landslides that have caused human impact over those that have not. **Figure 3.9D** illustrates that landslide records both from the NLD and newspapers rarely contain information about the size of landslides. Where this information was available, it was generally quoted as a weight in tonnes. Some articles would state the size of a landslide qualitatively (e.g., 'small' or 'large'), but we did not use these classifications on the grounds that landslide size varies by many orders of magnitude (Stark and Hovius, 2001; Malamud *et al.*, 2004a), and truly larger landslides are very rarely seen in Great Britain. Thus, a 'large' landslide to a British journalist may represent a relatively small landslide based on globally observed frequency-size statistics, and even in other British regions might be considered 'medium' or 'small'.

3.4.4 Assessing the Precision of Information Found Using Nexis UK

The precision to which each landslide event can be dated and located from newspaper information was estimated for all additional landslides identified from the Nexis UK archive. Spatial precision (S) is expressed in metres as a radius from the point location (of a given landslide event) given in the database. Date precision (D) is expressed as the amount of time either side of the date given in the database in which the landslide could have occurred. This is generally recorded in categories with increasing units of time (day, week, month, quarter, year). **Figure 3.10** shows frequency-size plots for the spatial and temporal precision respectively. Approximately 30 % of landslide events already existing in the NLD include an estimate of the spatial precision (**Figure 3.10A)**. Results are reasonably similar for the 2006 and 2012 periods. In both cases the locational precision of landslide events from the Nexis UK archive is slightly poorer than those landslides already existing in the NLD; in the NLD, the spatial precision peaks at a 100 m radius from the point location of a landslide event, whereas for the Nexis UK, the spatial precision peaks at a 1000 m radius. The date precision (D) of additional landslides identified from the Nexis UK archive is generally good (**Figure 3.10B**), with many landslides dated to within one day of occurrence and 65−75 % of landslides dated to within a week of occurrence. We hypothesise that this

is attributable to a generally short lag between event occurrence and reporting (whilst the event is of public interest).

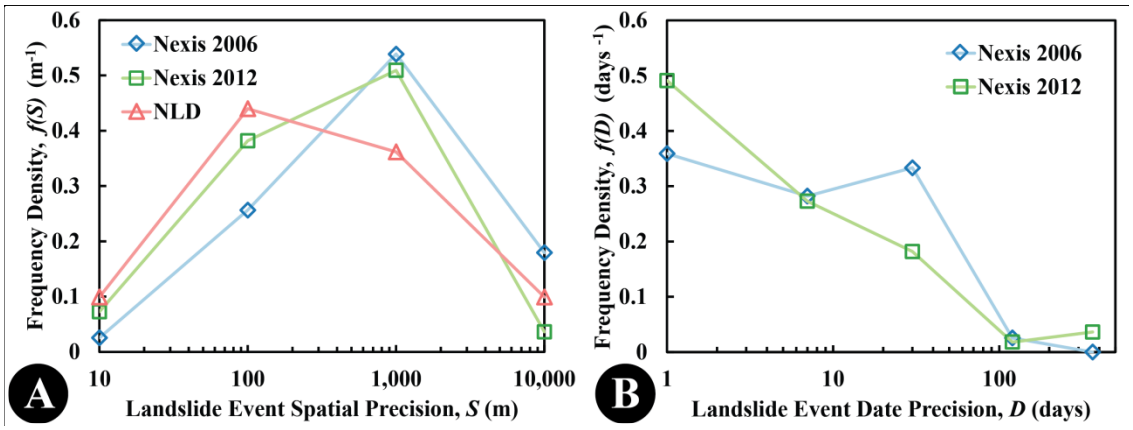


Figure 3.10 *Frequency density plots of precision of landslide information available from Nexis UK newspaper articles, compared to landslides that already exist in the NLD that occurred between 2006 and 2012 (where data exists). (**A**) Frequency density of spatial precision of landslide (N, E) location (S, defined as a radius surrounding that point in which the landslide is estimated to have occurred). Estimates of spatial precision are also available for approximately 30 % of entries in the NLD, shown in pink. (**B**) Frequency density f(D) of temporal precision of calendar date estimated to be when that landslide occurred (D) measured in days. This estimate of D is not included in existing NLD entries.*

In **Figure 3.11**, boxplots were used to show the time lag in weeks between a landslide event occurring (estimated from articles) and being reported in Nexis UK newspaper articles, classified by the dating precision of that landslide (see caption of **Figure 3.11**) for details). For landslides where dating precision was within one day, the median time lag between the event and reporting is 2 days. For landslides dated within a week, month and quarter, the median lag is equal to one unit of that time period. For landslides identified in both newspapers and the NLD, an estimate of the date precision is not available, but the median time lag for all these events was 2 days.

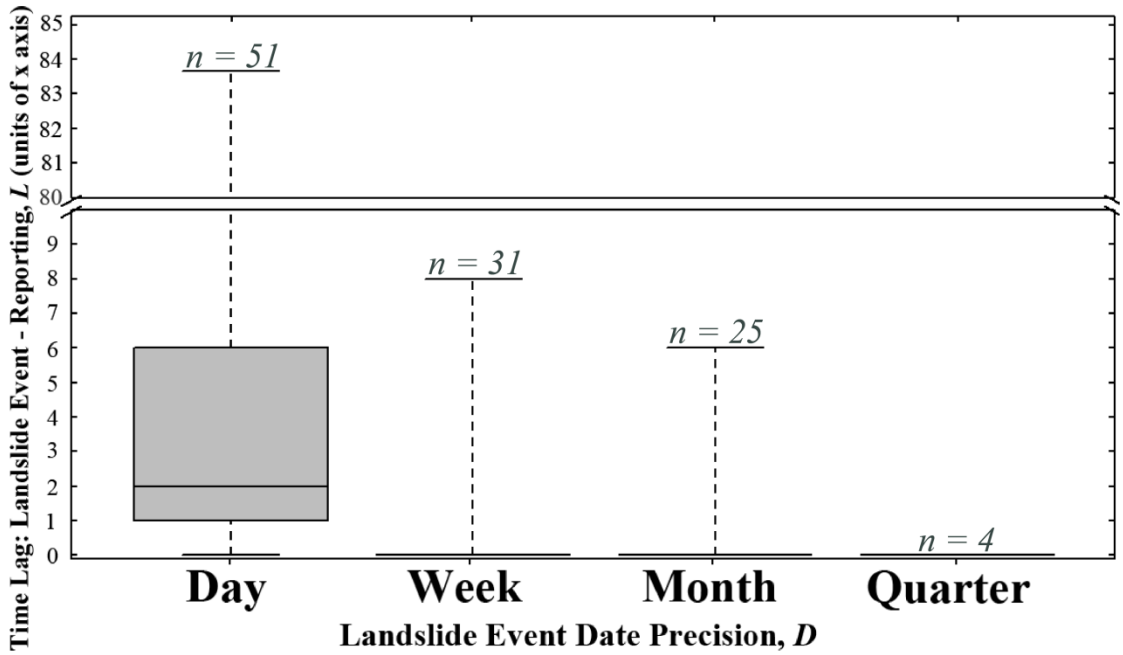


Figure 3.11 *Boxplot of time lag between when landslide is estimated to have occurred and when it was reported (L) (based on article publication date) for the 111 landslide events identified from the Nexis UK archive in 2006 and 2012. Lag is separated by the estimate of temporal precision (D) for each landslide and expressed in units of x. For example, if a landslide can be dated to within one day of occurrence, the time lag is measured in number of days between occurrence and publication. Similarly, if a landslide can be dated to within a week of occurrence, the time lag is expressed in number of weeks between occurrence and publication. Boxplot whiskers represent the full range of the data for each category.*

3.5 Discussion

In this **Chapter 3**, we have demonstrated that searching digital newspaper archives is an effective and robust method for adding richness to the NLD. In particular, the search methods we developed were consistently successful in:

(i) *Adding previously unrecorded landslide events* to the NLD for all but one month of the 24 months analysed (**Section 3.4.2, Figure 3.6**).

(ii) *Adding further confidence* to many of the existing landslide entries in the NLD by adding additional sources of information (**Section 3.4.3, Figure 3.2**).

(iii) *Augmenting the information recorded* for landslides in the NLD, particularly about their impact (**Section 3.4.3, Figure 3.9**).

With this proof of concept test, it should now be possible to apply our method to enrich NLD records of historic landslides occurring throughout the period covered by Nexis UK. Moving forward, our search terms could also be applied to supplement the existing sources of information used to alert of BGS of landslide events. This would provide the BGS with a relatively rapid method of 'reconnaissance' to guide whether further investigation (e.g., contact with council/land owner, site visit, remote sensing) may be required.

The most successful element of this work was the addition of landslide events to the NLD. This has resulted in a 122 % increase (for 2006) and 40 % increase (for 2012) in the total number of new landslide events recorded in the NLD. The spatial and temporal distribution, types and triggers for these additional landslides recorded using this method are consistent with existing understandings of landslide susceptibility in Great Britain. They also agree broadly with those already recorded in the NLD, which by definition is a 'patchwork' of methods and efforts devoted to data collection strategies (Foster *et al.*, 2012). This agreement provides a basis for added confidence in the NLD as a representative sample of contemporary landsliding in Great Britain, which looks to be growing more complete over time. No single resource will ever provide a complete record of recent landslide events, as events in rural or coastal areas with no impacts are likely to stay unreported, but this research reassures and enhances the current spatiotemporal record. The increasing proportion of events recorded in the NLD relative to those identified from the Nexis UK search highlights the influence of evolving data search–and-capture methodologies. Access to more social media resources, systematic processing and the adaptions of rules regarding the addition of smaller and engineered slope failures has greatly enhanced the 'live' recording of events (Pennington *et al.*, 2015).

Beyond this immediate application to enriching the NLD, this research has wider aims. By outlining in detail a clear methodology for developing and applying Boolean operators for searching digital archives of text data, we have provided earth scientists with a guide for exploiting the new sources of data about earth system processes opened up by the 'digital humanities' and projects like the British Newspaper Archive, which is scanning the vast holdings of historic newspapers held by the British Library to make them available for online searching (British Library, 2014). Following the systematic approach we have described in this **Chapter 3**, it should be possible to develop terms for searching these and other digital archives in order to (i) enrich the records of historic landslides held in the NLD and other landslide inventories (ii) develop similar databases for other hazards.

As with any method, there are uncertainties and biases involved in using such an approach, which we discuss in **Section 3.5.1** along with ways of overcoming them. In **Section 3.5.2** we then discuss how the bias towards events impacting humans could actually be useful in providing a rich source of data for quantifying the costs and other societal impacts of landsliding. In **Section 3.5.3** we go into more detail on how others might extend this research by applying to longer time periods and adopting a more automated approach.

3.5.1 Uncertainties and Biases Related to the Method

While searching newspaper archives offers an effective, relatively low cost method for gathering additional data about landslides and other natural hazard events, there are inevitably uncertainties and limitations to be considered. First, it requires subjective expert judgment to translate journalistic text into the data fields of the NLD. Sometimes relevant information is not explicitly within the news article, but can be inferred, and such inferences can vary between operators (Devoli *et al.,* 2007). In our case, we explicitly used two different people to search the Nexis UK regional newspapers and a one-day training period was performed to ensure consistent interpretation of results. Such 'investigator triangulation' is a well-established method for ensuring the robustness of qualitative research in social science (Baxter and Eyles, 1997).

Second, there are also systematic biases in media coverage that affect its use as a source of landslide inventory data. Media coverage tends to focus attention on large or 'novel' events and those with human interest (Moeller, 2006; Allan *et al.*, 2013) such as an impact on society (e.g., in the UK, road diversions, rail delays, homes being demolished or the closure of coastal footpaths). Also, while landslide events are relatively unusual and therefore generally newsworthy, media attention depends on perceptions of salience and if a small landslide occurs on the same day as a large election, the landslide may go unreported, whereas in a period of major landslide impacts (as observed in Great Britain in 2012), landslides may rank high in public interest and receive proportionally more coverage due to an availability bias (Pennington and Harrison, 2013). Thus, although the search strategy used here is systematic, the database we are searching is not a spatially or temporally homogeneous record of events.

3.5.2 Obtaining Information about Landslide Impact from Newspapers

By their very nature, newspaper articles primarily report on "landslides with consequences" (Guzzetti *et al.,* 2003, p. 472). In a major review of news coverage of disaster events, Quarantelli (1996) found that individual newspapers tend to report on average 90 stories about a particular event, and are most active in the post-event period, providing analytical coverage and a rich source of information about impacts. In **Figure 3.9C**, we showed that just over 50 % of landslide events in the NLD from 2006 onwards contain some information about impact, whereas 60 – 90 % of landslide events identified from the Nexis UK archive contained impact information. Moreover, we found examples of longitudinal

reporting of impacts, such as one newspaper article at the time of the event and another article a few months later reporting the remediation works undertaken.

One challenge in compiling records of landslide impacts is defining categories by which it can be measured. For example, Guzzetti (2000) and Petley (2012) use measures of the number of annual fatalities caused by landslides, Klose *et al.* (2014) put forward a methodology for measuring the impacts of landslides in economic terms, and Guzzetti *et al.* (2003) quantify the impact at a regional scale on population, transportation and properties. Schuster and Highland (2003) also note that very few studies consider the impact of landslides upon natural, non-human environments. Because of these difficulties and discrepancies in recording past events, there are few examples in the literature of robust, large-scale forecasting of the impacts of landslides.

Due to the original design and intended research purposes of the NLD, the existing categories in the NLD for recording the impacts of landslides were found to be somewhat insufficient for capturing the rich variety of information available in newspaper articles (see **Section 3.2.2** for a description of categories). While there are fields for number of fatalities, number of persons injured and cost, other impact information is largely recorded as free text. After analysis of Nexis UK articles from 2012 was complete and additional events and information added to the NLD, the list of impact information (for both landslides already existing in the NLD and additions from Nexis UK work) was organised into broad categories, which provide a first indication of the main types of impact observed in Great Britain in a particularly severe year. **Figure 3.12** shows an infographic of the principal types of impact observed – although it has been noted that the majority of landslides that occurred in 2012 were small shallow failures and in the coming years there may be different types of impact caused by larger, deep seated landslides that have a longer lag time between rainfall and triggering. Nonetheless, this impact information from 2012 now provides a baseline for comparison to other hazard impact data recording structures (see de Groeve *et al.* 2013 for a recent review).

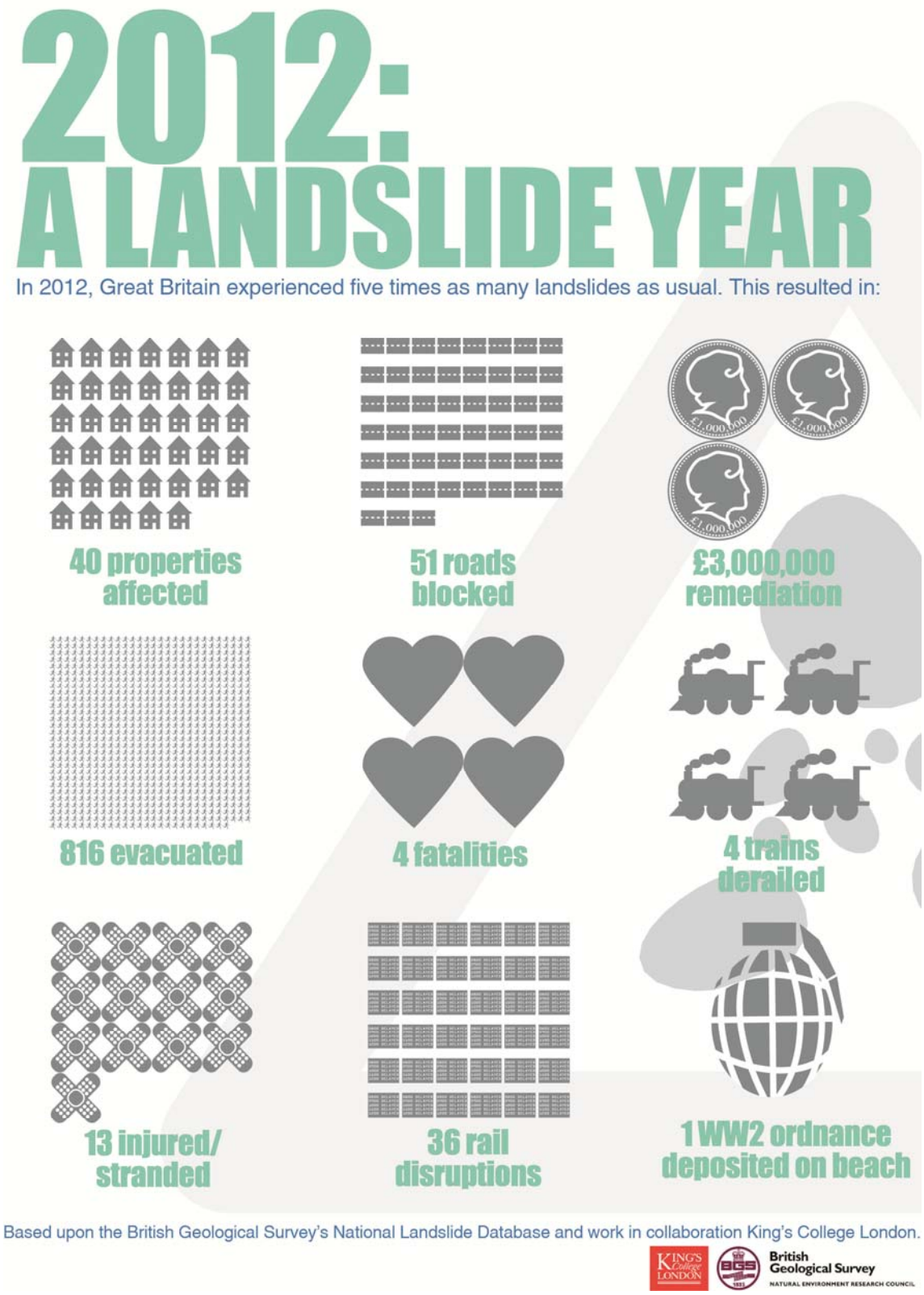


Figure 3.12 *Infographic of the main types of impacts caused by landslides in Great Britain in 2012. Data from landslide events in the NLD and additional events added in from searching the Nexis UK regional newspaper archive. The final category (WW2 ordnance deposited on beach) represents other more irregular or chaotic impacts.*

An ongoing piece of work with the BGS has been to map types of landslide impact on to the categories used by the Met Office Weather Observations Website (WOW) (Met Office,

2011). This could potentially result in restructuring how landslide impact is recorded in the NLD, and work towards a more standardised way of recording impact from different hazards. A more integrated approach between different hazard impact recording systems could potentially be beneficial when considering multiple interacting hazards (Gill and Malamud, 2014) in the UK wide hazard impact model. The Met Office WOW impact recording system records by the following criteria (Met Office, 2011):

- *Type of hazard* causing the impact (e.g., rain, wind, snow, combination, etc.).
- *What the hazard has impacted* (e.g., transport, utilities, property, leisure, combination, etc.).
- *Scale of the impact* on a likert scale range 1 – 4 (1 being light impact, 4 being severe impact).

Although still in the discussion stage, we have found that many of the likert scales in the Met Office WOW system are also relevant to landslide impact (an example for recording travel disruption is shown in **Table 3.3**), although it may be useful to add additional fields of information to build a more detailed picture of the impact. For example, when measuring travel disruption (**Table 3.3**), it may be useful to split this into modes of transport (e.g., road, rail, other), hierarchy (e.g., motorways, main roads, minor roads) and include binary categories for impacts more specific to landslides (e.g., whether a vehicle has physically been struck by a landslide). This could potentially lead to large numbers of redundant fields for most landslide events, but also create a much more useable dataset for considering landslide impact. Certainly, further work will need to strike a balance between level of detail in recording and being a user-friendly, manageable database that is widely used and easy to update.

Table 3.3 *Example of how travel disruption is recorded in the Met Office WOW system (Met Office, 2011)*

Level	Type of Impact
1	Difficult travelling conditions resulting in slow moving traffic, minor delays to commuting and public transport.
2	Longer journey times, local disruption to travel and service delays to public transport, fords impassable.
3	Roads closed, widespread disruption to public transport and/or Police advice not to travel.
4	Entire public transport networks suspended, prolonged road closures and/or motorists stranded in vehicles.

Although there is clearly potential to further mine newspaper articles for information about landslide impacts, there are biases such as overestimations, selective coverage and errors in interpretation of impact that must be taken into account (Freudenburg *et al.* 1996; Quarantelli, 1996). Typically, this would be countered by using the statements made from a range of articles. Such 'source triangulation' is well accepted in the social sciences for dealing with these problems (Baxter and Eyles, 1997). However, due to their local nature, we found that 65 % of landslide events were reported in only one article and where the event appeared in multiple articles, the information contained was often repeated verbatim. Nevertheless, newspaper reports can act as a near real-time alert that an impact has occurred and may need to be further investigated (Petrucci *et al.,* 2010).

3.5.3 Potential Extensions to the Method

As already described above, we are not the first to use newspaper articles as a source of information about landslide events. Newspaper have also been successfully drawn on as a major source of information about historical events (e.g., Guzzetti *et al.*, 1994; Elliott and Kirschbaum, 2007; Petley, 2012) and to supplement other landslide inventories (e.g., Miller *et al.*, 2009; Pradhan and Lee, 2010). Although these studies have undoubtedly been performed with attention to detail and in a systematic way, there is relatively little discussion within the literature of the detailed process of constructing a robust search strategy with the aim of capturing as many relevant articles as possible. It is hoped that by detailing the methodological steps involved and addressing related issues of uncertainty, this **Chapter 3** will make it easier for others to apply this method. We now discuss three potential extensions to the method we have explored in this **Chapter 3**: (i) extend archival searching farther back in time, (ii) increase speed and automation of the archival searching, (iii) extend archival searching method for landslides to other countries or other hazard databases.

(i) Extend archival searching farther back in time

To produce high quality landslide susceptibility maps and broadly have a good understanding of the landscape setting in which landslides occur across a region, we often require multi-temporal inventories of landslides, extending back over a number of decades. This is an issue for retrospective studies, as many landslides (particularly smaller ones) are 'erased' from the landscape via erosional processes within a few months to years (Malamud *et al.*, 2004a; Bell *et al.*, 2012). Thus, to produce historical inventories, we often rely on records of landslides from proxy sources. Indeed, in perhaps the best example of a long-

term (~90 years) archive of landslide events (The Italian AVI Project), over 60 % of records of landslide events come from newspapers, and the others from reports and interviews (Guzzetti *et al.*, 1999). Other examples include a database of historical landslides occurring in Utah from 1850 to 1978 (Elliott and Kirschbaum, 2007) and landslides occurring before 1990 in Nicaragua (Devoli *et al.*, 2007). Although the Nexis UK archive only extends back to 1998, there have been many advances in the digitisation, character recognition and compilation of historical UK newspaper sources going back considerably further, suggesting this method could be applied to much longer time periods to gain a better long-term understanding of landslide phenomena. For example, the British library has been undertaking a project to digitise its archive of newspapers extending back to 1800 (British Library, 2014). It is likely that the search terms listed in **Section 3.3.1** would need to be adjusted to take into account historical variations in terminology, but this presents an opportunity to gain further insight into landsliding in Great Britain over a relatively long timescale.

(ii) Increase speed and automation of the archival searching

There have been considerable developments in the field of automated newspaper content analysis using computers to identify the meaning of sentences within a text and extract information into a database; and this has been applied to fields such as political science (van Atteveldt *et al.*, 2008; Hopkins and King, 2009), economics (Sprenger and Welpe, 2011) and the policy dimensions of environmental phenomena such as Hurricanes (Soroka *et al.*, 2009) and climate change (Kirilenko and Stepchenkova, 2012). This could be of use to more rapidly process the large number of articles returned and retrospectively populate the database over longer time periods, particularly in countries where a large number of landslides occur annually. There are questions, however, about how easily this automated approach could be adapted to the creation of landslide event databases due to the indirect descriptions of events and the need for additional research to extract information (Discussed in **Sections 3.5.1** and **3.5.2**). There have been considerable advances in the ability to automate searches of large volumes of social media, so it is possible that now robust search terms have been developed, it may be possible to apply a more automated approach to the task.

(iii) Extend archival searching method for landslides to other countries or other hazard databases

The issues of database completeness are not specific to the field of landslides in Great Britain. As mentioned in **Section 3.2.1**, Van Den Eeckhaut *et al.* (2012) found that the majority of European countries that maintain landslide databases estimate the completeness to be around 50 %. At a global level, Guzzetti *et al.* (2012) estimated that only around 1 % of slopes have associated landslide inventory maps. Yet, detailed, systematic, well-produced landslide inventories are fundamental in both applied risk analysis (e.g., Harp *et al.*, 2011) and scientific research (e.g., Malamud *et al.*, 2004a). Indeed, it is acknowledged across many hazard-related disciplines that database incompleteness is an issue, and various proxy records have the potential to fill some of the gaps in our knowledge. Examples include Stucchi *et al.* (2004) for seismology, Barredo, (2007) for flooding and Blackford and Chambers (1991) with respect to climatology.

The method outlined in this **Chapter 3** has demonstrated a good ability to identify small landslides that might otherwise be missed by other methods of inventory production, historical landslides that may have been erased from the landscape and more generally, detailed accounts of hazard impact. The search terms outlined in **Step A5, Section 3.3.1** could be applied 'as is' to the remaining years of the Nexis UK archive (1998 to present), and perhaps with some further verification of temporal variations in terminology to the British Newspaper archive, which dates back to the 1800s (British Library, 2014). The Nexis archive also contains material from many countries across the globe, and has a similar level of coverage for France, Germany and the Netherlands (LexisNexis Academic, 2014). By clearly outlining the steps involved in search terminology experimentation (**Figure 3.1**), this method can now be applied broadly to other countries or other hazards to create robust, systematic inventories of hazard information from newspaper articles.

3.5.4 Implications for Research Performed in this Thesis

Chapters 2 and **3** of this thesis have addressed the topic of landslide inventories and databases, particularly focusing on the concept of 'completeness', which has important implications for modelling risk. This **Chapter 2** has shown ways in which a database can be supplemented to create a richer record of landsliding, but we have emphasised the fact that using archive methods will never result in a complete database, and potentially result in a bias towards 'landslides with consequences'.

The following **Chapters 4–6** represent a different direction from the research presented in this **Chapter 3**, as we shift our attention to the analysis of statistical patterns of size, shape and clustering of landslides. To perform this type of statistical analysis, substantially complete inventories are required, so that we can be sure we are looking at close to the full 'population' of landslide behaviour.

The findings of this **Chapter 3** become pertinent again in **Chapter 7** where we show the development of a Landslide-Road Impact Model (LRIM) which simulates triggered landslide events and their impact upon road networks. Although Great Britain is not particularly susceptible to medium or large triggered landslide events (e.g., hundreds or thousands of landslides occurring within a few minutes to weeks of an individual event such as an earthquake or a hurricane), this **Chapter 3** has shown the scale of impact that can be caused by a small landslide if it lands on a weak link in a transportation network. Indeed, **Figure 3.12** showed that there were at least 87 transport disruptions caused by 250 landslides in the UK in 2012, causing wide scale disruption that we have not been able to fully quantify. This also emphasises the difficulty in extracting observations of road network impact caused by landslides, even in a country with relatively good systems of reporting and recording data. This provides further justification for the LRIM (**Chapter 7**), which can be fairly rapidly applied to regions which are data poor, but also highlights the difficulty in obtaining quality datasets with which to confront the LRIM model output. This issue of data availability and quality is critically discussed in **Chapter 7**.

3.6 Summary and Conclusions

- This **Chapter 3** has set out a method to construct a set of Boolean terms and systematically search the Nexis UK archive of 568 regional newspapers for information about landslide events in Great Britain.
- When applied to all newspaper articles published in 2006 and 2012, this method added richness to the existing National Landslide Database (NLD) in three ways:
  (i) *Additional landslide events were added* that had not previously been recorded in the NLD, resulting in a 120 % and 40 % increase in the number of documented landslides in Great Britain in 2006 and 2012 respectively
  (ii) *The NLD records of landslide events were augmented, by adding more fields of information* and also providing additional sources of confirmation to many events, thus increasing the robustness of the database

(iii) *Landslide impact information could be obtained* from newspaper reports. There are some issues with uncertainty and inhomogeneities in media coverage of hazard events, which require caution.

- This method should be considered as supplementary to more robust methods of landslide database production (such as field investigation and remote sensing).
- This method represents a relatively quick, low-cost way of identifying events that may require further investigation.
- In explicitly outlining the steps involved in creating a robust, systematic search, we hope this method can be applied to other landslide and other hazard databases (such as flooding) to increase the richness of past records.

PART II: STATISTICAL MODELS OF LANDSLIDE SIZE, SHAPE AND CLUSTERING

Chapter 4: A Review of Statistical Models for Landslide Area

Chapter 4 Summary

This chapter presents a review of basic statistics and the literature concerned with statistical models for the frequency distribution of landslide areas. We later use these statistical models for the approach used in the Landslide Road Impact Model (**Chapter 7**) whereby we randomly select landslide areas from probability density functions in order to simulate triggered landslide events. In this chapter, we first introduce why and how we look at the frequency-area distribution of landslides, showing that by calculating the probability densities of individual landslide areas in a landslide inventory, we can compare the behaviour of different landslide inventories with different numbers of observations and estimate the probability of a landslide of a given area occurring. We then describe the overall behaviour of landslide areas in landslide inventories, showing that both an inverse Gamma or double Pareto probability density function appear to model well the distribution of landslide area probability densities in landslide inventories in many regions of medium- to high-topography that are susceptible to low mobility triggered landslide events. Both the inverse Gamma and double Pareto distributions model that as landslide area increases, the probability of observing a landslide of that area increases up until a 'rollover' point after which as landslide area increases, the probability of observing a landslide of a given size decreases as an inverse power-law. In an approximately chronological order, we review some of the main explanations proposed for this observed behaviour including: (i) censoring of small landslide areas, (ii) self-organised criticality, (iii) statistical patterns in topographic structure, (iv) soil strength and depth properties and (vi) anthropic activity. Although there is not consensus on why this statistical behaviour of landslide area is observed, the same patterns have been observed in increasing numbers of landslide inventories. We discuss whether the processes resulting in this behaviour may have implications for landslide shape and clustering, which we research in **Chapters 5** and **6**. We also discuss how we can use this generally observed behaviour to create synthetic triggered landslide event inventory maps to simulate the impact of trigged landslide events on road networks in the development of a Landslide Road Impact Model in **Chapter 7**.

*List of abbreviations used in **Chapter 4***

Abbreviation	Meaning	Section Introduced
DEM	Digital Elevation Model	4.4
pdf	Probability density function	4.2
SD	Standard Deviation	4.2
SOC	Self-Organised Criticality	4.4

List of variables used in ***Chapter 4***

Variable	Meaning	Section Introduced
α	Scale parameter for double Pareto pdf	4.3
β	Parameter that controls the power-law gradient of the left hand tail for double Pareto pdf	4.3
Γ	Gamma function ($\Gamma(x) = (x-1)!$) Where ! = factorial	4.3
δ	Bin width (right most minus left most area boundaries of the bin)	4.2
μ	Location parameter (mean) for normal and log-normal pdf	4.2
ρ	Scale parameter for inverse Gamma pdf	4.3
σ	Standard deviation parameter for normal and log-normal pdf	4.2
a	Shape parameter for inverse Gamma pdf	4.3
A_L	Individual landslide area	4.1
c	Cut-off' parameter controlling the minimum value for double Pareto pdf	4.3
$f(x)$	Frequency density	4.2
k	Power-law constant for power-law distributions	4.3
m	Cut-off' parameter controlling the maximum value for double Pareto pdf	4.3
M_w	Earthquake moment magnitude	4.3
n	Number of observations in a bin	4.2
N_{LS}	Number of landslides	4.2
$p(x)$	Probability density	4.2
s	Location parameter for inverse Gamma pdf	4.3
t	Parameter that controls the location of the maximum probability for double Pareto pdf	4.3

Equations used in ***Chapter 4***

Equation Number	Equation	Section Introduced
4.1	Calculation of frequency density	4.2
4.2	Calculation of probability density	4.2
4.3	Integral of probability density function	4.2
4.4	Gaussian probability density function	4.2
4.5	Log-normal probability density function	4.2
4.6	Inverse power-law probability density function	4.3
4.7	inverse Gamma probability density function	4.3
4.8	double Pareto probability density function	4.3

4.1 Introduction

This **Chapter 4** is an introduction to frequency-size statistics and a literature review which provides a basis for the following **Chapters 5** and **6** which present active research on statistical analysis of triggered landslide events. The motivation for attempting to statistically characterise triggered landslide events is to use generally applicable statistical models of landslide area, shape and clustering in the Landslide Road Impact Model (LRIM) (which will be introduced in **Chapter 7**) to simulate triggered landslide event inventories for a variety of locations, based on their general behaviour. Statistical models of landslide area have already been established in the literature, and we review them here. In the following **Chapters 5** and **6**, we use the techniques and approaches discussed here to develop statistical models of landslide shape and spatial clustering and test whether this behaviour is 'generally applicable'.

Hazard modelling research is driven by the questions of 'when', 'where' and 'how big' (Hochrainer, 2006), thus an important element of landslide hazard assessment is to understand the behaviour of landslide size (Guzzetti *et al.*, 2005; Guzzetti *et al.,* 2006a). Here, we refer to 'size' as the two-dimensional planar area of a landslide, although it is also common to conceptualise 'size' as the three-dimensional volume of a landslide (e.g., Brunetti *et al.*, 2009; Guzzetti *et al.*, 2009). Investigating the behaviour of landslide size is commonly done by examining the individual areas of landslide polygons in an inventory and comparing the number of small, medium and large landslides areas using frequency and probability densities (Brunetti *et al.*, 2009).

Over the last ten to twenty years, there has been considerable debate within the literature about (i) whether the frequency-area distribution of landslide areas (from various types of inventories) follows a 'general' distribution, (ii) what family of statistical distributions describes this behaviour and (iii) what causes it. Although there appears to be increasing agreement regarding the first of these points, there are very few examples in the literature of exploiting this generally applicable statistical behaviour of landslide areas to simulate landslide inventories. This approach has been used in probabilistic forecasting of other types of hazard (e.g., McGuire, 1995 for earthquakes; Apel *et al.,* 2006 for flooding), so it is one of the major aims of this thesis to apply this statistical modelling approach to simulate triggered landslide inventories and their impact upon regional road networks. In this section, we introduce the concept of landslide area probability distributions and some of the reasons why this behaviour might be observed.

In the following sections of **Chapter 4** we:

4.2 Briefly explain the basic concepts of frequency and probability density.

4.3 Review the literature discussing observed behaviour of landslide area frequency-size statistics.

4.4 Summarise potential physical explanations for observed landslide area frequency-size behaviour.

4.5 Discuss the implications of findings in **Sections 4.2 – 4.4** for research on other statistical behaviour of landslide inventories performed in **Chapters 5** and **6** and development of the Landslide Road Impact Model (LRIM) (**Chapter 7**).

4.6 Summarise and conclude the chapter.

Throughout this **Chapter 4**, we will use qualitative descriptions of individual landslide areas (A_L) in our discussions. **Table 4.1** gives approximate landslide area (A_L) ranges that these qualitative descriptions are based upon. The distinction between the categories in **Table 4.1** is based on the probability density distribution of landslide areas which will be explained in **Section 4.3**.

Table 4.1 *Qualitative descriptions of individual landslide area (A_L) sizes and the approximate range of values of individual landslide area these correspond to.*

Landslide size qualitative description	Landslide area, A_L
Very small	$A_L < 100$ m^2
Small	100 m^2 $< A_L \leq$ 400 m^2
Medium	400 m^2 $< A_L \leq$ 10,000 m^2
Large	10,000 m^2 $< A_L \leq$ 100,000 m^2
Very large	$A_L >$ 100,000 m^2

4.2 Brief Explanation of Frequency and Probability Density

4.2.1 Raw Histograms (Frequency)

Examining the frequency or probability density of any phenomena allows us to understand the relative proportions of observations of a given magnitude in a sample (i.e., what is the ratio of the number of small to medium to large observations) (Alcántara-Ayala, 2002). The first stage in calculation of frequency and probability density is to categorise the observation sizes into 'bins' of increasing magnitude. Using the example of landslide areas (A_L), these bins may be regularly spaced in linear coordinates (e.g., $0 < A_L \leq 100$ m^2, $100 < A_L \leq 200$ m^2, …), but when the phenomena spans several orders of magnitude, as is often the case for natural hazards, it is more practical to use variable bin sizes to avoid

under or over-weighting portions of the data due to the number of observations in that bin. In the case of landslides, it is convenient to increase the bin size logarithmically, as there are many more small to medium landslides than large to very large landslides (Stark and Hovius, 2001; Guzzetti *et al.,* 2002; Malamud *et al.,* 2004a Guzzetti, 2006a). This data can then be plotted as a 'raw' histogram (**Figure 4.1B**) showing the number-area distribution, with the 'raw' number of values n in a bin given as a function of each bin. This is demonstrated in **Figure 4.1** where **(A)** landslide polygons are separated into classes (by colour) depending on the area (A_L) of that polygon using approximately logarithmically increasing bin sizes and **(B)** the raw counts of number of landslides n in each bin is shown on a histogram.

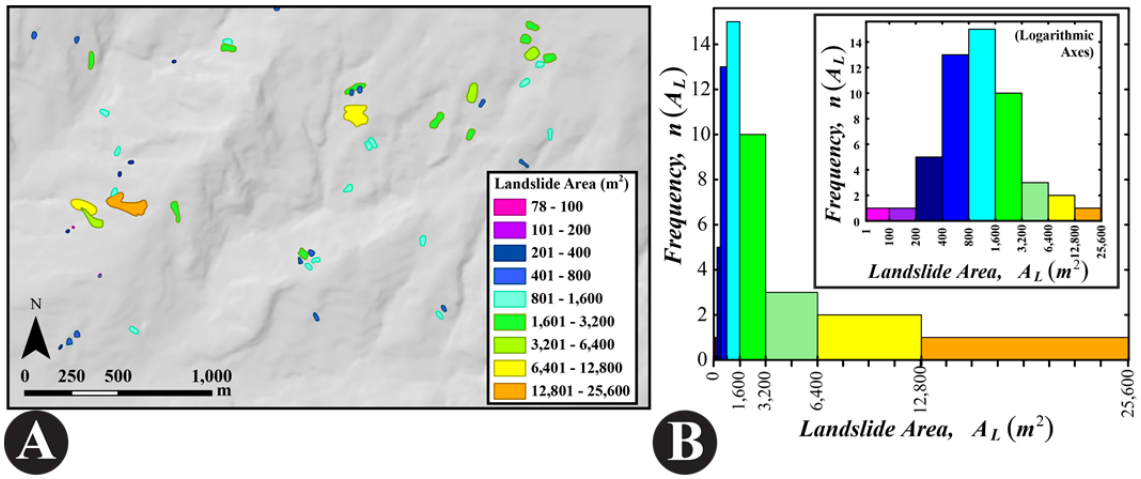


Figure 4.1 *Individual landslide area polygons as part of a triggered landslide event inventory in Collazzone, Central Italy in 1997 (Data: Cardinali et al. 2000).* ***(A)*** *Subset of inventory where 49 landslide polygons are plotted and colour varies with landslide area.* ***(B)*** *Histogram of landslide area counts using approximately logarithmically increasing bin sizes for landslides shown in (A) on linear axes and logarithmic x-axis (inset).*

Figure 4.1A demonstrates that landslide areas (A_L) range quite considerably over a relatively small region: from A_L = 78 m^2 to A_L = 13,200 m^2. **Figure 4.1B** demonstrates the raw counts of number of landslides in each area category on both a linear and logarithmic x-axis (inset). We see that the majority of landslide areas (43 out of 49) are between 200 m^2 < A_L ≤ 3,200 m^2. We also observe that it is much easier to visualise the distribution of landslide areas using a logarithmic x-axis due to landslide area varying by three orders of magnitude.

4.2.2 Frequency Density

As mentioned in **Section 4.2.1**, bin sizes can be varied. However, by varying the width of the bin, one can artificially increase or decrease the number of values in that bin (taken to

an extreme). Therefore, so that all counts are equivalent to unit size bins, the number of values in a bin (n) is normalized by the bin width (δ) to give the frequency density $f(x)$:

$$f(x) = \frac{n}{\delta} \quad (4.1)$$

Where:
n = number of observations in a given bin
δ = bin width (right most minus left most area boundaries of the bin)

The frequency density effectively normalizes the number of values in that bin to 'unit' size bins, i.e., how many values are there in a given bin if that bin were unit sized (Turcotte and Malamud, 2004). For example, in **Figure 4.1B**, there are n = 5 landslide areas in the bin category 200 < A_L < 400 m^2, giving a frequency density of $f(A_L) = n/\delta$ = 5/(400 m^2 – 200 m^2) = 5/(200 m^2) = 0.025 m^{-2}. An example of landslide area frequency density for two triggered landslide event inventories is shown in **Figure 4.2A**.

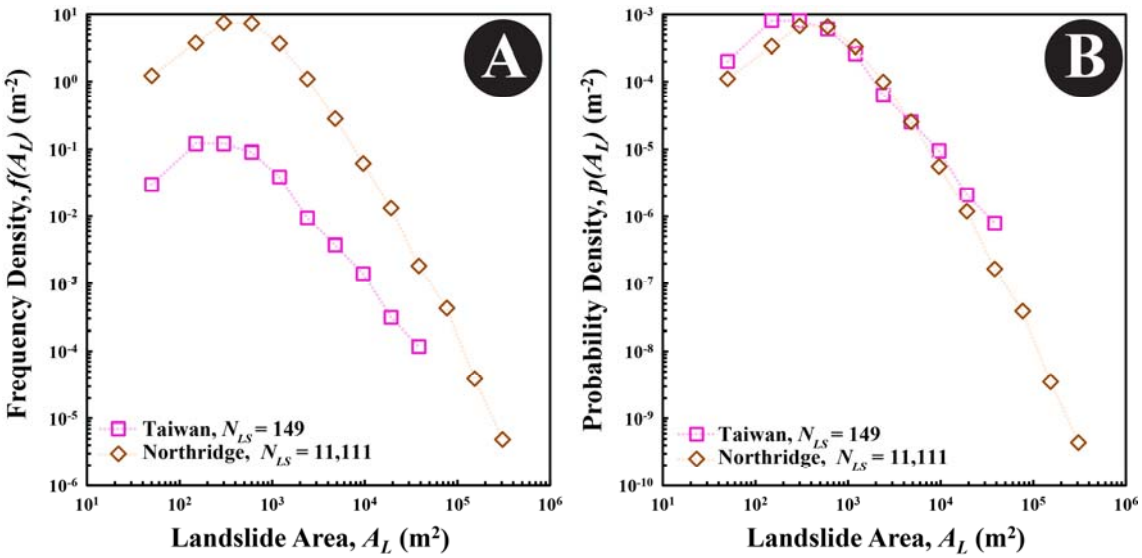


Figure 4.2 *Calculation of* ***(A)*** *Frequency Density $f(A_L)$ and* ***(B)*** *Probability Density $p(A_L)$ of landslide area for two triggered landslide event inventories. Those inventories are: N_{LS} = 149 landslides triggered by Typhoon Megi in Taiwan, 2009 (Chen, personal communication, 2014) and N_{LS} = 11,111 landslides triggered by the 1994 Northridge Earthquake in the USA (Harp and Jibson, 1995). Landslide areas are binned using approximately logarithmically increasing bin sizes. These inventories will be described in detail in* ***Chapters 5–7*** *of this thesis.*

4.2.3 Probability Density

Probability density is calculated by dividing the frequency density $f(x)$ (**Eq. 4.1**) by the total number of observations (N_{LS}) (e.g., the total number of landslides in the inventory), giving the probability density, $p(x)$:

$$p(x) = \frac{1}{N_{LS}} \times \frac{n}{\delta} \quad (4.2)$$

Where:
N_{LS} = *total number of observations in entire inventory*
n = *number of observations in a given bin*
δ = *bin width (right most minus left most area boundaries of the bin)*

By normalizing by the total number of observations (N_{LS}), different datasets can be compared on the same scale (e.g., the distribution of landslide areas in a landslide inventory of 100 landslides can be compared to landslide areas in a landslide inventory of 10,000 landslides) (Cha, 2007). An example is shown in **Figure 4.2** where the probability density distributions of areas in two landslide inventories are compared. One inventory has a relatively small number of observations (N_{LS} = 149 landslides triggered by Typhoon Megi in Taiwan, 2009) and one is has a relatively large number of observations (N_{LS} = 11,111 landslides triggered by the Northridge Earthquake, USA, 1994).

Using frequency densities in **Figure 4.2A**, the two orders of magnitude disparity in total number of observations between the two inventories is visible by the vertical offset. Whereas, in **Figure 4.2B**, the probability densities lie almost on top of one another because the total area under each curve has been effectively normalized to 1.0 (see following **Section 4.2.4**), and the very small probability densities for the larger landslide areas on the far right for Northridge contribute very little to the overall integration of the curve to 1.0. Because in **Figure 4.2B** the curves are now normalized by the total number of landslides in each inventory, it is easier to compare the relative probability of observing a landslide of a given size in each inventory. Calculation of the probability of observing a landslide of a given area is useful for hazard and risk assessment (Picarelli *et al.*, 2005; Guzzetti *et al.*, 2006a).

4.2.4 Probability Density Functions

A probability density function (pdf) (i.e., a continuous function) can be fit to a given set of observed probability densities (e.g., **Figure 4.3**), so that the probability of observing a

landslide of any[1] given area can be calculated (Stark and Hovius, 2001; Guzzetti *et al.*, 2002; Malamud *et al.,* 2004a; Guzzetti *et al.,* 2006a). Although it is typical to visualise the probability density function with observed values of probability densities (as in **Figure 4.3**), the pdf should actually be fit to the cumulative densities, to avoid uncertainties related to how the probability densities were binned (White *et al.* 2008). The integral of a pdf (i.e., the area under the pdf) is by definition equal to one:

$$\int_{-\infty}^{\infty} p(x)dx = 1.0 \qquad (4.3)$$

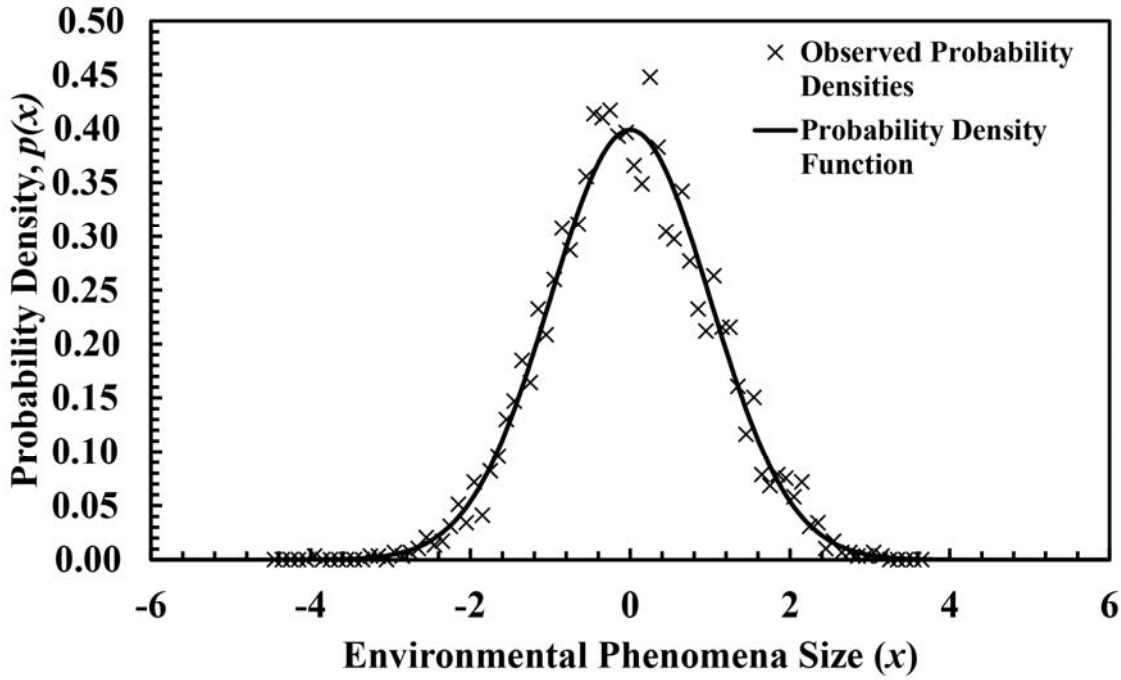


Figure 4.3 *Probability density distribution of synthetic environmental deviates (randomly generated from a Gaussian distribution). Shown are the probability densities using bins with equal width of $\delta = 0.1$ and the best-fit Gaussian probability density function (pdf) (where mean $\mu = 0.0$ and standard deviation $\sigma = 1.0$) (**Eq. 4.4**).*

[1] Throughout this chapter, the term '***any*** landslide area' refers to any landslide area within the upper and lower bounds of observations (allowing for reasonable extrapolation). It is not yet clear whether very extreme landslide events follow the same behaviour as observations are very rare, thus prediction far beyond the bounds of observed values must be exercised with caution (Sachs *et al.,* 2012).

For many environmental phenomena, it has been often assumed that the data follow a Gaussian (normal) or log-normal distribution (Esmen and Hammad, 1977; Limpert *et al.*, 2001; Newman, 2005). A Gaussian probability density function is given by:

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} \times e^{\frac{(x-\mu)^2}{2\sigma^2}} \quad (4.4)$$

Where:
μ= *mean of values* x
σ = *standard deviation of the values* x.

Another commonly used probability distribution for natural phenomena, which we show here because of its contrast to the Gaussian and because all values are positive, is the log-normal probability density function (Limpert *et al.,* 2001):

$$p(x) = \frac{1}{x\,\sigma\sqrt{2\pi}}\, exp\left(-\frac{1}{2\sigma^2}(\ln(x)-\mu)^2\right) \quad (4.5)$$

Where:
μ = *mean of natural logarithm of* x *variable (ln(x))*
σ = *standard deviation of natural logarithm of* x *variable (ln(x))*

Examples of both probability distributions are given in **Figure 4.4**.

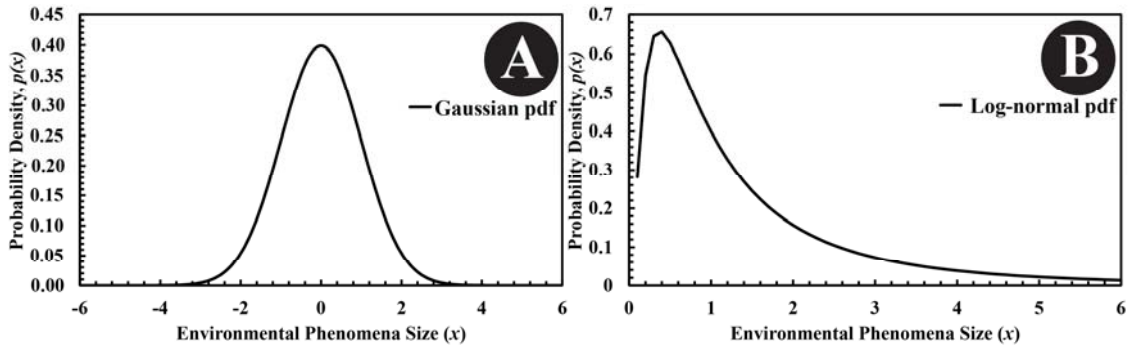


Figure 4.4 *Examples of probability density functions commonly used in the natural sciences:* ***(A)*** *Standard Normal (Gaussian) Probability Density Function with mean* μ *=0.0 and standard deviation* σ *=1.0 (**Eq. 4.4**)* ***(B)*** *Log-Normal Probability Density Function with* μ *=0.0 and* σ *=1.0 (**Eq. 4.5**).*

In a normal distribution (**Figure 4.4A**), the probability density function is symmetrical about the mean and decays exponentially away from the mean; about 95.4 % of the data are within ±2 S.D (two standard deviations) of the mean, thus the probability of 'extreme' events (in the case of landslides, very large or very small landslide areas far from the mean value) is relatively low (Kang and Wen, 2000). The log-normal distribution (**Figure 4.4B**) is skewed to the left in this case, and has a slightly heavier right-hand tail compared to the right-hand tail of the Gaussian distribution shown in **Figure 4.4A**. This means that there is

a greater probability of observing phenomena of a very large size (e.g., the largest landslide areas) than would be modelled by a Gaussian distribution. Thousands of other probability density functions exist to model the probability density distribution of data. These vary in terms of how they look (e.g., shape, skew), complexity (e.g., number of parameters required to describe the pdf) and how they model the overall behaviour of a system (e.g., how the tails of the distribution decay) (Parzen, 1962; Witt and Malamud, 2013). In the following **Section 4.3** we show that observed behaviour of landslide area probability density deviates significantly from a Gaussian or log-normal pdf, which has important implications for our understanding of landslide hazard and risk.

4.3 Observed Probability Density Patterns in Landslide Inventories

For various types of landslide inventory, many authors have found that the probability density distribution of landslide areas (A_L) does not follow a Gaussian or log-normal shape pdf (e.g., see review of 27 studies by Van Den Eeckhaut *et al.,* 2007). Moreover, observed landslide areas span many orders of magnitude: from very small landslides (of the order of 10 m^2), through to some of the largest landslides on the Earth's surface being tens of square kilometres in area, and submarine landslides up to two orders of magnitude larger than this (Guzzetti, 2006a; Korup *et al.,* 2007). The behaviour of landslide area probability density distributions appears to be similar across many inventories, the reasons and implications of which are discussed in **Sections 4.4** and **4.5**. In **Section 4.3.1**, we describe the observed statistical behaviour for medium and large landslide areas ($A_L > \approx 400$ m^2). In **Section 4.3.2** we discuss observed statistical behaviour for small landslide areas ($A_L < \approx 400$ m^2). In **Section 4.3.3** we discuss statistical distributions which have been used to describe the overall pattern of behaviour of landslide areas in various inventories.

4.3.1 Observed Patterns in the Probability Density Distribution of Large Landslide Areas

For historical, seasonal and triggered landslide event inventories (described in **Chapter 2**), there are typically many more small to medium landslide areas (A_L) than large to very large landslide areas. However, the number of large landslides (the right hand tail of the distribution) decays more slowly than modelled by a Gaussian or log-normal distribution. Several authors have put forward that the shape of the probability density distribution of medium to very large landslide areas is better modelled by an inverse power-law distribution (Hovius *et al.,* 1997; Pelletier *et al.,* 1997; Stark and Hovius, 2001; Guzzetti *et al.,*

2002; Crosta *et al.,* 2003; Guthrie and Evans, 2004; Malamud *et al.,* 2004a and others). An inverse power-law relationship is given by:

$$p(x) = x^{-k} \quad (4.6)$$

Where:
p(x) = probability density
x = observation size
k = power-law constant

In **Eq. 4.6**, the probability of a landslide of a given area (x_i) occurring is inversely proportional to the landslide area, with k the inverse power-law exponent, resulting in the characteristic straight line 'heavy tail' when plotted on logarithmic axes (**Figure 4.5**). Heavy tailed (sometimes referred to as 'fat' or 'long' tailed) in this case means that the right tail of the inverse power-law decays more slowly than the right hand 'tail' for a Gaussian distribution (an exponential) (Malamud, 2004). In practical terms, if landslide area probability density distributions follows an inverse power-law, this means we would expect proportionally more 'extreme' events (large landslides) than modelled by a Gaussian or log-normal distribution, and that landslide area spans many more orders of magnitude. This has important implications for hazard modelling and forecasting (Guzzetti *et al.,* 2005).

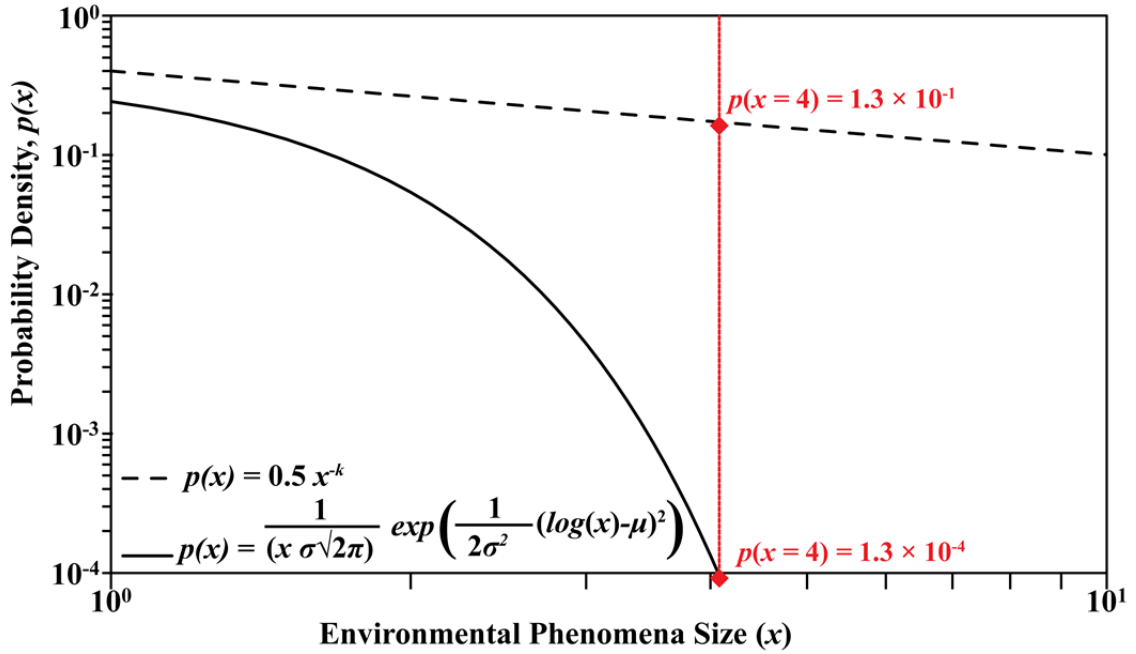


Figure 4.5 *Comparison between the right hand tails of two probability distribution functions: (i) exponential decay of a Gaussian distribution (**Eq. 4.4**) where σ=1 and μ = 0 and (ii) inverse power-law (**Eq. 4.6**) where k = -0.6. The axes are logarithmic. In this example, for an 'extreme' event x= 4, the probability of occurrence is two orders of magnitude higher when modelled by an inverse power-law distribution than an exponential distribution (marked on the figure with red diamonds). Note, it would be possible to show an exponential distribution where p(x) is larger for an exponential distribution than an inverse power-law, by changing the parameter values. This figure is included to demonstrate the 'shape' of the distributions more than exact values.*

4.3.2 Observed Patterns in the Probability Density Distribution of Small Area Landslides

Regarding the probability density distribution of small landslide areas, there is somewhat more conjecture in the literature regarding what probability density function best models this behaviour, and this is more dependent on the type of landslide inventory. This is because small landslides are more rapidly erased from the landscape by erosional processes, re-vegetation and anthropic activity and may be less easily discernible from imagery (Malamud *et al.,* 2004a; Guthrie and Evans, 2007; Bell *et al.*, 2012; Guzzetti *et al.,* 2012). It is also because in many landslide inventories, landslide areas below a given size (e.g., landslide areas A_L < ≈100 m^2) are often not recorded, as discussed in **Chapter 2**. Thus, when looking at historical and multi-temporal inventories, or triggered event inventories that are incomplete, it is likely that the 'true' probability density distribution of small landslide areas is not reflected in the inventory data, as proportionally more small landslides will be missing. However, when looking at a substantially complete triggered event inventory (as defined in **Chapter 2**), we can be more confident that we are sampling the entire distribution of landslide areas roughly proportionally, or at least down to the minimum

areas that the authors say have been systematically gathered. Although there is some debate about the exact shape of landslide area probability distributions, there is increasing agreement that for triggered events, very small (e.g., A_L < ≈100 m^2) landslides are very rare or not observable in the landscape, and then the probability density of small landslide areas increases until they reach a maximum probability of occurrence for landslide areas at about A_L = 400 m^2 – 1000 m^2 (Stark and Hovius, 2001; Malamud *et al.,* 2004a). In practical terms, this means we observe few very small landslide areas, some small landslide areas and many medium landslide areas. To illustrate this, **Figure 4.6** shows an example from an inventory of landslides triggered by rapid snowmelt in 1997 in the Collazzone region (Central Italy) (a subset of this **Figure 4.6** is shown in **Figure 4.1**).

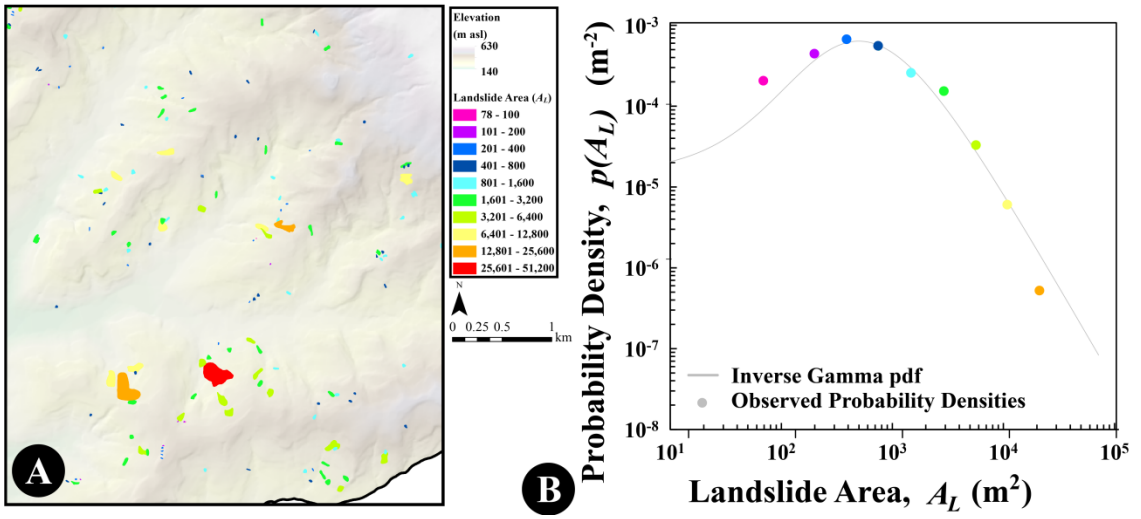


Figure 4.6 *Triggered landslide event inventory and corresponding probability densities. Inventory of 413 landslides triggered by rapid snowmelt in Collazzone, Central Italy, 1997 (subset of Umbria snowmelt inventory, described later in* ***Section 4.3.3*** *in* ***Table 4.2****) (Cardinali et al., 2000).* ***(A)*** *Subset of landslide inventory map where landslide polygon colour corresponds to area.* ***(B)*** *Probability densities of observed landslide areas (points) binned into approximately logarithmically increasing bin sizes and inverse Gamma probability density function (****Eq. 4.7****) fit to data (parameter values ρ = 1.4, a = 1280 m², s = −132 m²).*

In the example shown in **Figure 4.6A**, we observe relatively low probabilities of very small landslides (magenta coloured polygons), high-probabilities of medium sized landslides (purple, blue and green colours) and low probabilities of large landslides (yellows to red colours). This results in a rollover in probability density around A_L = 400 m^2 (**Figure 4.6B**). Shown in grey is the inverse Gamma pdf with parameter values from Malamud *et al.* (2004a) which we discuss in detail in the following **Section 4.3.3**.

4.3.3 Probability Density Functions fit to Landslide Areas

The probability density behaviour of low- mobility landslide areas has been described for multiple triggered landslide event inventories using an inverse Gamma pdf (Malamud *et al.,*

2004a) (**Eq. 4.7**) and a double Pareto pdf (Stark and Hovius, 2001) (**Eq. 4.8**). Low mobility is broadly defined as all types of landslide excluding the 'end member' forms of rockfalls and very long debris flows, where the run out zone may be the combination of several landslide source areas and the behaviour is more like a hyper concentrated stream flow (Malamud *et al.,* 2004a).

Both the inverse Gamma and double Pareto distributions have an inverse power-law decay for medium to large values of A_L, and a 'rollover' where the maximum probability density occurs at small to medium values of A_L. For small values of A_L, the shape of this part of the pdf (i.e., the rollover) has been described as an exponential by Malamud *et al.* (2004a) and a power-law by Stark and Hovius (2001). Depending on the completeness of the inventory examined, the parameter values describing the best fit to the data will vary. The equations for these two probability distributions are given below, where $p(A_L)$ is the probability of a landslide area (A_L) occurring:

Inverse Gamma probability density function (Malamud *et al.*, 2004):

$$p(A_L) = \frac{1}{a\Gamma(\rho)}\left[\frac{a}{A_L - s}\right]^{\rho+1} \exp\left[-\frac{a}{A_L - s}\right] \quad (4.7)$$

Where:
a = Location parameter primarily controlling the position of the maximum probability (rollover)
s = Shape parameter primarily controlling the gradient of the left hand tail of the distribution
ρ = Scale parameter primarily controlling the inverse power-law decay of the right hand tail

Double Pareto probability density function (Stark and Hovius, 2001):

$$p(A_L) = \left[\frac{\beta}{t}\left(1 - \left(\frac{1+\left(\frac{m}{t}\right)^{-\alpha}}{1+\left(\frac{c}{t}\right)^{-\alpha}}\right)^{\left(\frac{\beta}{\alpha}\right)}\right)^{-1}\right]\left[\frac{\left(1+\left(\frac{m}{t}\right)^{-\alpha}\right)^{\left(\frac{\beta}{\alpha}\right)}}{\left(1+\left(\frac{A_L}{t}\right)^{-\alpha}\right)^{\left(1+\frac{\beta}{\alpha}\right)}}\left(\frac{A_L}{t}\right)^{(-\alpha-1)}\right] \quad (4.8)$$

Where:
α = Scale parameter that controls the inverse power-law decay of the right hand tail
β = Parameter that controls the power-law gradient of the left hand tail
t = Parameter that controls the location of the maximum probability (rollover)
c = 'Cut-off' parameter controlling the minimum value of landslide area for which the pdf can be evaluated
m = 'Cut-off' parameter controlling the maximum value of landslide area for which the pdf can be evaluated

Theoretical examples of the inverse Gamma and double Pareto pdfs are shown in **Figure 4.7** on logarithmic axes – we observe that although the distributions differ slightly, there is agreement in how they describe the general behaviour of landslide area; initially probability

density increases with landslide area to an inflection point (the 'rollover') where the maximum probability occurs. After this point, probability density decays relatively gently until one reaches maximum observed landslide areas of the order of 1 km^2. It is not fully known whether this inverse power-law behaviour can be extrapolated for extremely large landslide areas (Stark and Guzzetti, 2009), due to a relative paucity of observations of recent events (i.e., in the last century) (Korup *et al.,* 2007), so some authors introduce a 'cut-off' of 1 km^2, above which the probability density function is not extrapolated (Stark and Hovius, 2001).

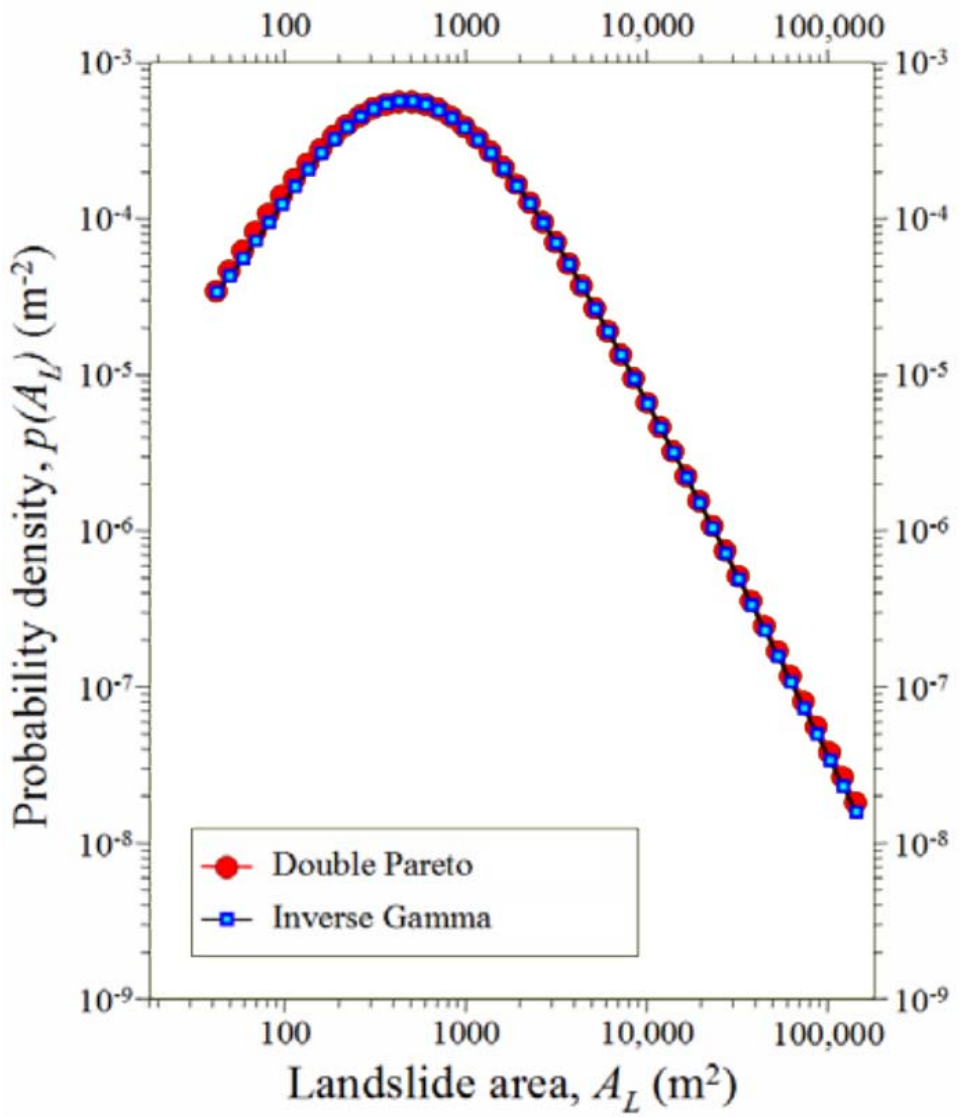


Figure 4.7 *Comparison of the inverse Gamma pdf and double Pareto pdfs fit to N_{LS} = 4,233 landslide areas triggered by snowmelt in Umbria in 1997 (inventory described in* ***Table 4.2****. Figure taken from Guzzetti (2006).*

Both the presence of the rollover at small to medium landslide areas and the landslide area at which it occurs have been debated within the literature. It would seem more intuitive that the entire distribution of landslide areas (from smallest to largest) would simply follow

an inverse power-law decay, as observed with the frequency-size distributions of many other hazards such as earthquake magnitude (Gutenberg and Richter, 1944), forest fire burned areas (Malamud *et al.*, 2005) and hailstone sizes (Cheng *et al.,* 1985). Thus, it was initially argued (Stark and Hovius, 2001; Brardinoni and Church, 2004, Hungr *et al.,* 2008) that the deviation from an inverse power-law decay for the probability of small landslide areas was the result of under-sampling of small landslide areas in the inventories used—either by small landslides being erased from the landscape via erosional processes in historical inventories or small landslides being below the minimum resolvable resolution of the imagery in multi-temporal or triggered event inventories. However, Malamud *et al.* (2004a) demonstrated that when using three substantially complete triggered landslide event inventories, where for each inventory there was substantial confidence that small landslide areas are not being systematically under sampled, the rollover is consistently observed at landslide areas significantly larger than the minimum resolvable area of the imagery, suggesting that the rollover is real and reflects mass-wasting process general behaviour.

The concept that a probability density function with a rollover for small landslide areas reflects process is somewhat backed up by the fact that the best-fit parameters of the pdf to observed data appear to be approximately the same for triggered landslide event inventories in medium-high topographic regions across the world, regardless of the triggering mechanism or geologic setting (Malamud *et al.,* 2004a). Malamud *et al.* (2004a) examined three large triggered landslide event inventories (described in **Table 4.2**) that can be considered as 'substantially complete' in terms of representing the full population of landslide areas proportionally (criteria for 'completeness' described in **Table 2.5**, **Chapter 2**). Although these inventories differ in terms of topographic setting, climate, landslide triggering mechanism and total number of landslides, Malamud *et al.* (2004a) found that the probability density distributions of all three inventories were similar and reasonably well modelled by an inverse Gamma probability density function with parameter values $\rho = 1.4$, $a = 1280$ m^2, $s = -132$ m^2 and that the location of the rollover consistently occurred at around 400m^2, shown in **Figure 4.8**.

Table 4.2 *Information about the three substantially complete triggered landslide event inventories found by Malamud et al. (2004a) to be reasonably well modelled by an inverse Gamma probability density function with the same parameter values for each location (**Figure 4.8**). Table adapted from Malamud et al. (2004a).*

	Northridge	**Guatemala**	**Umbria**
Location	California, USA	Guatemala	Umbria, Italy
Trigger	Northridge Earthquake (M_w 6.7), January 1994	Hurricane Mitch, October 1998	Rapid Snowmelt, January 1997
Inventory source	Harp and Jibson (1995)	Bucknam *et al.* (2001)	Cardinali *et al.* (2000)
Method	Mapped from 1:60,000 aerial photography taken hours after event + field surveying.	Mapped from 1:40,000 scale aerial photographs taken Jan-March 2000 + limited field surveying.	Mapped from 1: 20,000 aerial imagery flown within months of the event.
Number of landslides, N_{LS}	11,111	9,594	4,233
Total area of landslides, $\sum A_L$(km^2)	23.8	29.5	12.7
Largest Landslide Area, Max(A_L) (km^2)	0.259	3.87	0.156
Mean Landslide Area, A_L (km^2)	2.14×10^{-3}	3.07×10^{-3}	3.01×10^{-3}

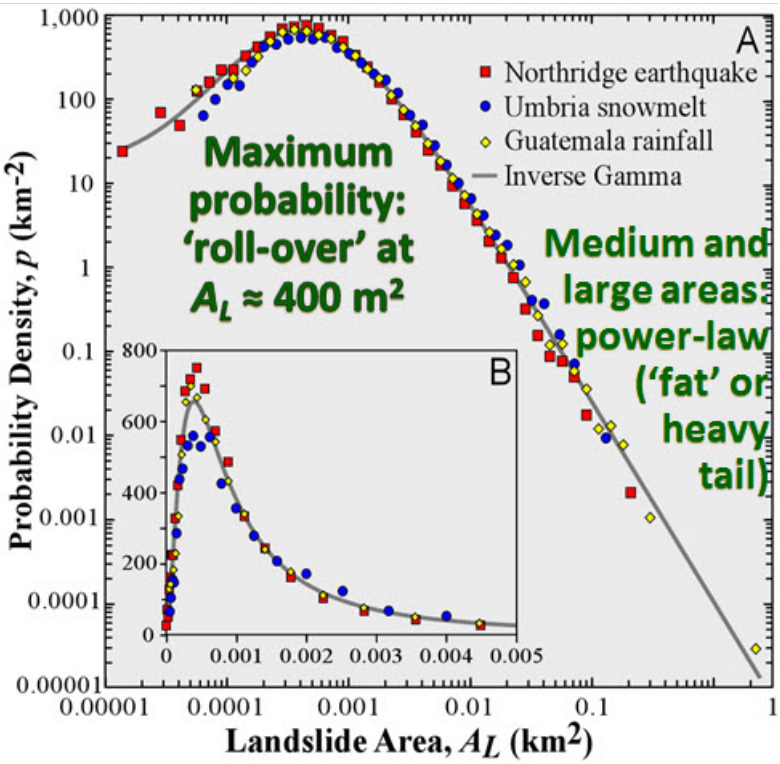


Figure 4.8 *The inverse Gamma probability density function fit to three substantially complete triggered landslide event inventories (Malamud et al., 2004a). Inventories described in* ***Table 4.2****. The inventories are:* N_{LS} *= 11,111 landslides triggered by the 1994 Northridge Earthquake in California, USA;* N_{LS} *= 9,594 landslides triggered by 1998 Hurricane Mitch in Guatemala and* N_{LS} *= 4,233 landslides triggered by 1997 snowmelt in Umbria. The inverse Gamma pdf (****Eq. 4.7****) fit to the observed landslide area probability densities has the parameter values:* ρ *= 1.4,* $a = 1.28 \times 10^{-3}$ *km² (1280 m²),* $s = -1.32 \times 10^{-4}$ *km² (–132 m²).*

Increasingly, authors are finding that other triggered landslide event inventories also follow the inverse Gamma or double Pareto pdf. For example, Chen (2009) for two Typhoon triggered landslide inventories in Taiwan; Fan *et al.* (2012) for landslides triggered by the 2008 Wenchuan Earthquake in China; Martha *et al.* (2015) for landslides triggered by rainfall in Uttarakhand, India in 2013. Others have shown that in inventories where some smaller landslides are censored (either historical inventories where small landslides have been erased from the landscape or seasonal/event inventories produced at a small scale, **Chapter 2**, **Section 2.3**), the observed distribution will follow a similar inverse power-law decay for medium to large landslides as shown in **Figure 4.8**, but the location of the maximum probability (rollover) will occur at a larger landslide area, if at all (Guzzetti *et al.,* 2002). This deviation from the 'generally observed' pdf can be used as an indicator of inventory completeness (Malamud, 2004a, Guzzetti, 2006; Ray *et al.,* 2009; Borgomeo *et al.*, 2014; Razak and Brinksma, 2014). The physical processes resulting in this 'general' behaviour are not fully understood, but key theories are discussed in **Section 4.4**.

4.4 Why do we observe this 'General' Behaviour in Landslide Area Probability Density Functions?

4.4.1 Issues with Data versus Physical Explanations

For more than four decades, authors have suggested that the frequency-size behaviour of large landslide areas follows an inverse power-law decay (e.g., Fuji, 1969; Whitehouse and Griffiths, 1983; Sugai *et al.,* 1994; Pelletier *et al.,* 1997). Although it was initially argued that the deviation from an inverse power-law decay for landslide areas was simply a result of censoring of small and medium landslide areas at the stage of landslide inventory production (e.g., Stark and Hovius, 2001; Brardinoni and Church, 2004; Hungr *et al.,* 2008), several authors have shown that the rollover of the pdf occurs at landslide areas considerably larger than the resolution the inventory was produced at (Pelletier *et al.,* 1997; Malamud *et al.,* 2004a; Frattini and Crosta, 2014). Thus the general inverse Gamma or double Pareto shape of the pdf fit to landslide areas in many inventories more likely reflects physical processes of mass wasting rather than simply methodological biases in the inventories used. Several viable theories have been put forward to explain this behaviour but consensus does not yet exist. In the following section we outline some of these theories, in a roughly chronological order in which they appear in the literature, as some theories build upon one another and other theories are in disagreement. The major 'themes' of these theories are:

4.4.2 Self Organised Criticality.

4.4.3 Topographic Structure.

4.4.4 Material and Depth Properties.

4.4.5 Anthropic Action.

4.4.2 Theory 1: Self-Organized Criticality

The inverse power-law behaviour exhibited in the probability density functions (pdfs) of medium to very large area landslides has been attributed to the concept of self-organised criticality (SOC). SOC is a concept that originated in the physics community to describe the complex, scale invariant (i.e., power-law) behaviour of many natural systems (Frigg, 2003), whereby small perturbations in a system may result in small changes or system-wide changes, and the frequency-size distribution of these changes follows a power-law (Clauset *et al.*, 2009).

The power-law frequency-size statistics of a self-organised critical system can be conceptualised by Bak *et al.* (1987)'s sand pile model (**Figure 4.9**) where sand grains are randomly 'dropped' over a finite grid. Each grid cell can hold up to four sand grains. As the number of grains in the cell increases, its stability decreases. When four grains is exceeded, the sand grains are distributed to the four adjacent cells in a sand 'avalanche' which may in turn cause a cascade of now unstable sand grains in neighbouring cells, potentially resulting in one large avalanche. The sand-pile model over time results in many 'small-sized' events (e.g., just one sand grain lost from the grid), some 'medium-sized' events (where the cascade results in some grains lost from the grid), and few 'large-sized' events where there are many sand grains lost from the grid in a cascade. In the sand pile model, the occurrence of large sand avalanches is dependent on the growth and coalescence of small unstable regions, and experimental studies have shown that the resultant probability density distribution of avalanche sizes (number of grains lost per model step) has been shown to have an inverse power-law decay with exponent $k = -1$ (Malamud and Turcotte, 1999).

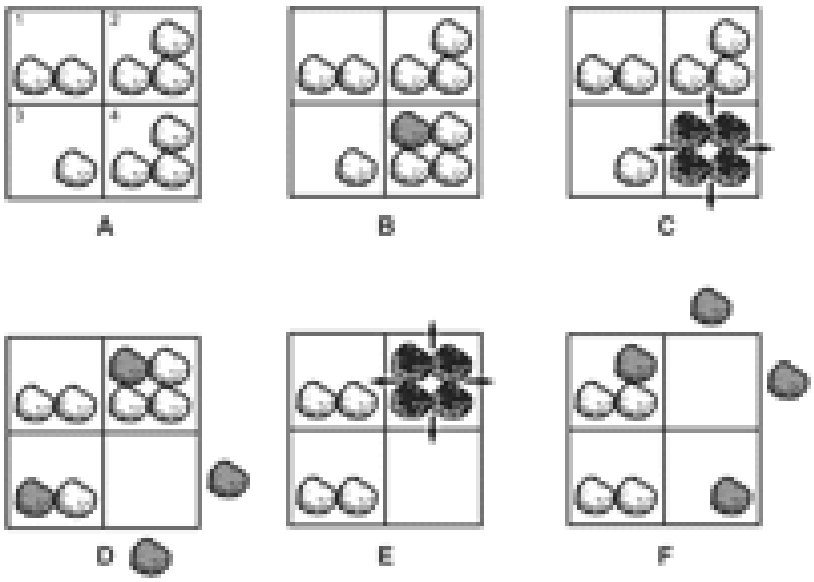


Figure 4.9 *Simple illustration of the sand pile model over four cells. Sand grains are randomly dropped over the grid; when a cell contains four sand grains, it becomes unstable and causes an 'avalanche' where the grains are redistributed to neighbouring cells. Stages A–F demonstrate the state of the model at different time steps. At stage C, the bottom right cell results in an avalanche. At stage D/E the previous avalanche has caused another cell to become unstable, resulting in a second avalanche demonstrating how unstable regions can coalesce. Image from Turcotte (2006).*

There is an instinctive similarity between avalanches of sand grains and real-world landslides. Coupled with the inverse power-law behaviour of the frequency-size distributions of both the sand-pile model avalanches and observed landslide areas in inventories, several authors (listed in Pelletier *et al.*, 1997) put forward that SOC could be used to explain the observed inverse power-law frequency-size distributions of landslide

areas (Hergarten, 2003). However, other than the instinctive similarity and similar resultant statistical behaviour, the physical explanation linking SOC to landslide frequency-size behaviour is somewhat weak.

There are several issues with this theory of self-organized criticality causing this behaviour of landslide areas: (i) the deviation from the inverse power-law for small values of A_L, which has been shown to be real (Guzzetti *et al.,* 2002; Malamud *et al.*, 2004a, 2004b), (ii) the exponent of the best fit inverse power-law is considerably larger for real landslides than theoretical sand avalanches (e.g., sand-pile model) meaning larger landslides/avalanches are more frequent in theoretical sand pile models than observed in inventories (Malamud *et al.*, 2004b) and (iii) the principles driving the model (the constant input of grains of sand randomly dropped over a region) do not directly relate (or 'map on') to the processes of tectonic uplift and fluvial incision which drive gravity driven mass movements in reality (Hergarten, 2003).

Authors such as Densmore *et al.* (1998); Piegari *et al.* (2009) and Lehmann and Or (2012) have continued to develop and 'tune' the basic sand pile model to more accurately reflect observed landslide probability density distributions by incorporating factors such as time-dependent weakening, whereby stability of a cell is governed by both the number of grains in a cell and the time since the last avalanche within that cell (Densmore *et al.,* 1998). Nonetheless, authors continue to question how accurately these models reflect the true physics governing these systems (Pelletier *et al.,* 1997; Guzzetti *et al.,* 2002; Hergarten, 2003, Malamud *et al.,* 2004b).

4.4.3 Theory 2: Topographic Structure

Pelletier *et al.* (1997) argued that the sand pile model is only dependent upon the continuous random input of sand grains to the model, and that in reality, water plays a major role in triggering landslides. Pelletier *et al.* (1997) showed that the frequency size statistics of patches of soil with soil moisture content greater than certain thresholds followed an inverse power-law decay, caused by small scale variations in evapotranspiration. This suggests that the probability density distribution of landslide areas is dependent on a complex spatial pattern of locations of steep slopes and patches of high soil moisture. To explain the flattening of the inverse power-law for small landslide areas, Pelletier *et al.* (1997) proposed that at large scales, the landscape is broken up by drainage divides, whereas at small scales, topography is effectively smooth and fails as one unit.

Thus there are more 'medium' sized failures of these entire smooth units, resulting in less multiple very small failures within these smooth units.

Guthrie *et al.* (2008) built upon Pelletier *et al.* (1997)'s theories about ridges and divides limiting the size of landslides. They created a cellular automata model to test the hypothesis that the majority of landslides initiate in the mid to upper portion of slopes, entrain more material as they travel down slope and flow until they reach a physiographic constraint such as a river channel or area of shallower terrain. Resultantly, large landslides are only limited by the maximum slope size whereas the number of small landslides is limited by the slope length, although this was only tested with debris flows.

This concept of topographic surface being a major controlling factor on landslide frequency-size distributions was revisited by Frattini and Crosta (2013). Using high resolution DEMs, they abstracted the surface topography to a series of triangles of varying size, where each triangle represents a 'smooth' topographic unit. The frequency-size distribution of these triangular tiles follows an inverse power-law with a rollover at smaller scales. The exponent of the inverse power-law is higher for shallower slopes as it is possible to have larger units of smooth slope in flatter areas. Assuming the maximum landslide area is defined by the size of the triangular topographic unit, Frattini and Crosta (2013) hypothesized that the combination of availability of triangular units of a given size and the probability of failure within each unit (defined by slope and material) defines the frequency-size behaviour of landslide areas observed in reality.

4.4.4 Material and Depth Properties

Katz and Aharonov (2006) built upon Pelletier *et al.* (1997)'s ideas of small areas of homogeneous land failing as a whole, thus reducing the number of very small landslides and introducing a 'characteristic' scale of landslide areas which reflects the idea of a rollover in the pdf observed in real triggered event inventories. These theories were tested by creating a physical model of a slope made of sand where landslides are artificially triggered by vibrations in various directions. They observed that small landslides were generally shallow failures in the homogeneous relatively weak sediments close to the surface of a slope where the area of the landslide will be maximized and only limited by the extent (i.e., depth) of the heterogeneous material. Larger landslides occurred as the result of fractures and discontinuities which are heterogeneously distributed in the sub-surface of the slope. This is somewhat supported by Hurst *et al.* (2013)'s analysis of the frequency-size distribution of landslides in Great Britain, separated out by lithological unit. Hurst *et al.*

(2013) found smaller landslides occurred less frequently in the harder, more consolidated units of rock (which could be interpreted as more subjected to fracturing), whereas the rollover was more present in landslides that had occurred in superficial or weakly consolidated lithological units.

Stark and Guzzetti (2009) built upon Katz and Aharonov (2006)'s theory, hypothesising that the primary control over the frequency-size distribution of landslide areas is variance in cohesion; which is low for shallow soil where small landslides occur, but greater for bed rock where large landslides occur. In Stark and Guzzetti (2009)'s theoretical/stochastic model of individual landslide rupture and movement, they stated that a landslide will occur when the weight of the material exceeds its shear strength, which is a product of cohesion and friction. In shallow unconstrained soils, where cohesion is relatively week, the landslides will tend to continue to grow until cohesion around the perimeter of the landslide is significant enough to abate further movement. Hence a 'characteristic' landslide area around the rollover is caused by relatively few shallow landslides being constrained from growing. After this characteristic area, landslide area then becomes constrained by increasing cohesion with depth and that larger landslide areas tend to deeper and occur in bedrock, and thus the inverse power-law probability density distribution is dominated by frictional strength (Stark and Guzzetti 2009). Other authors such as Milledge *et al.* (2014) have built upon these concepts and introduced factors of landslide geometry and root cohesion as limiting factors in the minimum area of failure. Frattini and Crosta (2013) also found that cohesion limits minimum possible landslide area (although they argued that topographic surface and slope angle played a greater part in defining landslide frequency-size distributions). Indeed Brunetti *et al.* (2014) found that in an inventory of landslides on Mars, there was no rollover for small to medium sized landslides, which they attributed to a lack of cohesive material limiting minimum landslide area to fail, which supports the findings of Stark and Guzzetti (2009).

More recently, Katz *et al.* (2014) presented results of numerical simulations of landslide sliding plane initiation and propagation, showing that amount of material available to fail (and thus, landslide size) is a function of slope strength and angle. In their model, stronger materials (i.e., higher cohesion) result in a failure plane closer to the slope surface which does not propagate up to the slope crest, thus the amount of material that will fail is smaller. Moreover, the steeper the slope, the greater the amount of material available to fail. Extrapolating these simulations to real life, Katz *et al.* (2014) suggest that the failure

plane of most landslides occurs at the transition between unconsolidated materials and bedrock, and so landslide characteristic size (i.e., the rollover) is essentially limited by depth of unconsolidated materials, which only varies by a few meters globally. They explained the largest landslides are not the result of failure in this superficial layer but failures in deeper pre-existing fractures. Although Chen *et al.* (2015) have recently published a response to Katz *et al.* (2014) showing that in their model the depth of landslides decreased with increasing slope angle, which is in direct opposition to previous findings.

4.4.5 Anthropic Action

Van Den Eeckhaut *et al.* (2007) proposed that in regions of more gentle topography, not susceptible to 'extreme' triggered landslide events, the frequency-size distribution of observed landslides was more dominated by human activity. They argued that large landslides were generally caused by 'natural' triggers, and do not commonly occur in the present era, thus are under sampled in existing inventories. Whereas, human activity triggers many more small landslides.

4.4.6 Summary of Physical Explanations for Observed Landslide Frequency Size Statistics

Assuming that the observed probability density distribution of landslide area does indeed follow either a 'generally applicable' inverse Gamma or double Pareto probability density function (i.e., has a rollover for small to medium landslide areas and inverse power-law decay for large landslide areas), **Section 4.4** has outlined the evolution of theories to explain why this behaviour is observed. An increasing number of authors have started to use deviation from the established inverse Gamma or double Pareto pdfs as an indicator of inventory incompleteness, adding weight to the argument that the behaviour is real and not simply a result of under sampling of small landslide areas. The physical explanations for the observed rollover and inverse power-law decay tend to focus on (i) properties of the landscape that are inverse power-law distributed, such as soil moisture or topography and (ii) soil characteristics related to depth, cohesion and friction, and authors have been able to reproduce observed landslide area frequency size statistics by modelling from both approaches. Broadly, it appears that the hypothesis of cohesion and friction being the major driving factors behind observed landslide frequency-size distributions appears to be the gathering traction in the scientific community (e.g., Stark and Guzzetti, 2009; Frattini and Crosta, 2013, Milledge *et al.,* 2014; Katz *et al.* 2014), although is far from agreed upon.

4.5 Implications of Landslide Area Frequency-Size Statistics for Research Performed in this Thesis

It is perhaps the decade-long debate about the driving factors and potential methodological biases that has resulted in there being relatively few examples of actually applying these statistical patterns for civil protection purposes. For example, Malamud *et al.* (2004b) illustrated potential applications of these general frequency-size statistics for forecasting total landslide surface area affected during a triggered landslide event, and estimation of erosion. Some of the authors of the Malamud *et al.* (2004b) paper on the 'generally applicable' nature of the inverse Gamma pdf to model triggered event landslide area demonstrated how these frequency-size statistics could be applied to give a probabilistic estimate of landslide of a given area occurring to improve landslide hazard assessment at the basin scale (Guzzetti *et al.* 2005). Yet, perhaps because it is more typical to use more deterministic methods, these statistical patterns have not been widely used by other authors, as the statistical patterns cannot be easily modelled without first understanding the physical processes that cause them.

In this **Chapter 4**, we have shown that there is evidence by some authors that low-mobility landslide areas in triggered landslide event inventories follow a 'generally observed' statistical distribution with a rollover for small to medium landslide areas and inverse power-law decay for medium to very large landslide areas. We believe that both the double Pareto (Stark and Hovius, 2001) and the inverse Gamma (Malamud *et al.*, 2004a) probability density function are reasonable models as they exert similar overall behaviour, although we now mainly focus on the inverse Gamma pdf as a model of triggered landslide event area behaviour, where 'generally applicable' parameter values are published (Malamud *et al.* 2004a). We have presented a range of potential reasons why the behaviour in landslide area might be similar for many locations across the globe, although there is yet to be consensus within the literature. However, the range of potential reasons reviewed here set an important basis for the research performed in the following **Chapters 5** and **6**, where we investigate whether these physical processes may also result in general patterns of landslide shape and spatial clustering. Then, in **Chapter 7**, we show the development of a statistical model that uses the 'generally applicable' behaviour of landslide area in an applied sense to create 'synthetic' triggered landslide event inventory maps and overlay these with maps of regional road networks. These synthetic inventory maps are created by randomly selecting landslide areas from the inverse Gamma pdf and landslide shapes from statistical distributions established in **Chapter 5**, and this process is repeated hundreds of times in a

Monte-Carlo type simulation. Thus, although we are not certain of all of the deterministic processes that result in the real distribution of landslide shapes and sizes in triggered event inventories, we can recreate synthetic versions of these maps based on their general statistical behaviour.

4.6 Summary and Conclusions

In this **Chapter 4**, we have:

- *Introduced how and why we look at hazards in the frequency-size domain*:
    - Showing that understanding the proportion of small to medium to large size hazards has important implications for forecasting risk.
    - Showed how probability densities are calculated, allowing the comparison between events with different numbers of observations.
    - Showed how fitting a continuous probability density function to a set of observations allows the calculation of the probability of observing a hazard of any given size.
- *Described the observed behaviour of landslide area* probability density:
    - For medium to very large values of landslide area, the probability distribution tends to follow an inverse power-law decay. This means there are more large landslide areas than modelled by a Gaussian distribution.
    - The probabilities of small landslide area tend to deviate from this inverse power-law decay, and exhibit a rollover at around 400 m^2. This means that there is a high probability of observing a small landslide area (e.g., at 400 m^2), but a lower probability of observing smaller landslide areas (< 400 m^2).
    - This behaviour has been observed in various triggered landslide event inventories which differ in terms of topographic setting and triggering event (e.g., earthquake, heavy rainfall), thus it is thought that the probability density distribution of landslide areas is 'generally applicable.
- *Reviewed the literature looking at why we observe this behaviour* in landslide area probability density distribution:
    - It was initially thought that the deviation from the inverse power-law for the probability of very small landslide areas was simply a result of under sampling small landslides.

- It has been shown that this rollover (maximum probability) is real and occurs at a finer resolution than the resolution inventories are created.
- There is a lack of consensus in the literature about the physical processes governing this 'generally applicable' behaviour but key theories include (i) properties of the landscape that are inverse power-law distributed, such as soil moisture or topography and (ii) soil characteristics related to depth, cohesion and friction.

- Although the processes are not fully understood, we accept that landslide area does follow a generally observed probability density function, which we can now use to create synthetic triggered landslide event inventory maps to simulate the impact of triggered landslide events on road networks in **Chapter 7** of this thesis.

Chapter 5: What Shape is a Landslide? Statistical Distributions of Landslide Length-to-width Ratio

Chapter 5 Summary

This primary research chapter presents a methodology to quantify the shape of landslides by measuring landslide length (L) to width (W) ratio (L/W), and investigate the variability in L/W. We use two large, substantially complete triggered-event inventories available publically: 11,111 earthquake-triggered landslides (1994 Northridge, USA) and 9,594 rainfall triggered landslides (1998 Hurricane Mitch, Guatemala). We experiment with four methods to abstract landslide polygons to ellipses and find the best method is obtained by fitting a convex hull (CH) to each landslide polygon, approximating an ellipse with the equivalent area (A_{CH}) and perimeter (P_{CH}) of the convex hull and then scaling this ellipse to match the area of the original landslide (A_L). The goodness-of-fit (e) of elliptical approximations was calculated using a measure of the area of intersection (A_I) between the original landslide polygon area (A_L) and the elliptical approximation. We find that > 80 % of landslides can be reasonably well approximated by an ellipse in both inventories. Length-to-width ratios (L/W) of ellipses are calculated for all landslides with elliptical goodness-of-fit $e \geq 0.5$. The statistical distributions of length-to-width (L/W) ratios are then examined, both for all landslide areas (A_L) and subcategories of landslide areas. Using Maximum-Likelihood Estimation bootstrapping techniques and Monte Carlo Kolmogorov-Smirnov testing, we find that an inverse Gamma probability density function (pdf) is a good statistical model for the probability density distribution of landslide L/W, with parameters of the inverse Gamma pdf depending on the landslide area category considered. The methodology is then applied to five additional 'lower confidence' triggered landslide event inventories with 149–851 landslides in each inventory, to confront whether landslide shape exhibits 'generally applicable behaviour'. There is some similarity in L/W behaviour in landslide inventories where the trigger involves water, where L/W typically increases with landslide area (A_L). For the one earthquake-triggered landslide inventory (Northridge), L/W tends to decrease with A_L. This suggests a potential 'semi-generally applicable' behaviour of landslide shape, depending on the triggering mechanism, although there is uncertainty due to the small sample sizes in the lower confidence inventories. The findings from this **Chapter 5** are used in the landslide-road impact model (LRIM, **Chapter 7**) to randomly select a length-to-width ratio from an appropriate pdf in order to define the dimensions of an ellipse for each modelled landslide.

List of abbreviations used in ***Chapter 5***

Abbreviation	Meaning	Section Introduced
CH	Convex Hull	5.5
E$f(A_{CH}, P_{CH})$	Quadratic ellipse (E) of convex hull (CH)	5.5
E$f(A_{CH}, P_{CH})\rightarrow(A_L)$	Quadratic ellipse (E) of convex hull (CH) scaled to landslide area (A_L)	5.5
E$f(A_L, P_L)$	Quadratic ellipse (E) of landslide (L)	5.5
E_{SD}	Standard deviation ellipse	5.5
K-S	Kolmogorov-Smirnov Test	5.6
LRIM	Landslide-road impact model	5.1
MLE	Maximum likelihood estimation	5.6
pdf	Probability density function	5.6
SD	Standard deviation	5.1

List of variables used in ***Chapter 5***

Variable	Meaning	Section Introduced
Γ	Gamma function ($\Gamma(x) = (x-1)!$) Where ! = factorial	5.6
ϵ	Number of decimal places precision for significance level of Monte-Carlo Kolmogorov-Smirnov goodness-of-fit test	5.6
ρ	Scale parameter for inverse Gamma pdf	5.6
a	Shape parameter for inverse Gamma pdf	5.6
A_{CH}	Area of convex hull (CH)	5.5
A_E	Area of ellipse (E)	5.5
A_I	Area of intersection (I) between landslide polygon and elliptical approximation	5.5
A_L	Area of landslide (L)	5.1
D	Kolmogorov-Smirnov test distance between two cumulative frequency curves	5.6
D_{OBS}	Kolmogorov-Smirnov test: distance between observed cumulative frequency (OBS) and MLE fit cumulative frequency curves	5.6
D_{SIM}	Kolmogorov-Smirnov test: distance between simulated data cumulative frequency (SIM) and MLE fit cumulative frequency curves	5.6
e	Ellipticity goodness-of-fit measure	5.5
$F(x)$	Theoretical cumulative distribution of x	5.6
$F_n(x)$	Empirical cumulative distribution of x	5.6
L	Length	5.3
L/W	Length-to-width ratio	5.2
M_W	Earthquake moment magnitude	5.4

Variable	Meaning	Section Introduced
n	Number or count (different variable types)	5.6
n_{iter}	Number of iterations for Monte-Carlo Kolmogorov-Smirnov goodness-of-fit test	5.6
N_{LS}	Number of landslides (LS)	5.3
p	Kolmogorov-Smirnov goodness-of-fit test significance level	5.6
P_{CH}	Perimeter of convex hull	5.5
P_E	Perimeter of ellipse (E)	5.5
P_L	Perimeter of landslide (L)	5.5
s	Location parameter for inverse Gamma pdf	5.6
Sup	Supremum (maximum)	5.6
W	Width	5.3

List of equations used in ***Chapter 5***

Equation Number	Equation	Section Introduced
5.1	Area of an ellipse	5.5
5.2	Approximation of perimeter of an ellipse	5.5
5.3	Calculation of ellipse length to width ratio from area and perimeter	5.5
5.4	Ellipticity goodness-of-fit measure	5.5
5.5	Inverse Gamma probability density function	5.6
5.6	Kolmogorov-Smirnov goodness-of-fit test	5.6
5.7	Number of iterations for Monte-Carlo Kolmogorov-Smirnov goodness-of-fit test	5.6

5.1 Introduction

In **Chapter 7,** we show the development of a landslide-road impact model (LRIM), which creates 'synthetic' triggered landslide event inventory maps by randomly selecting landslide areas (A_L) from already established landslide area probability density functions (introduced in the previous **Chapter 4**) and semi-randomly dropping these landslide areas over a study region that has been conditioned by a landslide susceptibility map. This process is repeated one hundred times in a Monte-Carlo type simulation to create one hundred different synthetic triggered landslide event inventory maps. In LRIM, each individual landslide area must be given a shape in order to be mapped as a polygon. We will show in **Section 5.3** that the shapes of landslides in real triggered landslide event inventories are typically

irregular and vary quite considerably. The Monte-Carlo approach (repeating a process hundreds of times) adopted in LRIM is quite computationally demanding, and thus we require a relatively 'simple' technique to give landslides a shape that also captures some of the observed variety in landslide shape. In order to model landslide shapes in the LRIM simply but semi-realistically, we perform primary research here to systematically quantify the shape of real landslides in triggered landslide event inventories, and then investigate the statistical variability in landslide shape to test whether we observe any 'generally applicable' behaviour of landslide shape.

We will do this by investigating the shape of landslides in two large, substantially complete triggered landslide event inventories, containing over 20,000 landslides in total (described in **Section 5.4**) and then five additional inventories which we consider to be 'lower confidence' and with an order of magnitude fewer observations (described in **Section 5.4**, **Table 5.2**). Because of the scale of the analysis performed here, it is not practical to individually measure the shape of each individual landslide manually. Thus in **Section 5.5**, we experiment with various methods to rapidly and systematically quantify the shapes of large numbers landslides in a semi-automated GIS procedure.

In addition to contributing to the LRIM modelling process, the research performed in this **Chapter 5** aims to contribute to our knowledge of the statistical behaviour of triggered landslide events that was introduced in **Chapter 4**, linking into the concepts of 'generally applicable' non-linear behaviour and the processes that might cause this behaviour.

In this **Chapter 5** we:

5.2 *Discuss what 'shape' means* spatially and how one can measure it.

5.3 *Review literature discussing what shape landslides typically are* and what processes may govern this.

5.4 *Introduce seven landslide inventories used* later in the chapter analyses (**Section 5.5** and **5.6)**.

5.5 *Describe four methods to abstract the shape of large numbers of landslide polygons* in triggered landslide event inventories to ellipses and test the goodness-of-fit of an elliptical approximation. We apply these methods to two large triggered landslide event inventories.

5.6 *Investigate how length-to-width ratio of landslide elliptical approximations varies* with landslide area and how to test this in a statistically robust manner. This is first applied to two large substantially complete triggered event inventories and then to five additional smaller, lower confidence inventories.

5.7 *Critically discuss results* and potential physical reasons for observed L/W behaviour and the implications of the findings of this Chapter.

5.8 *Summarise and conclude* **Chapter 5**.

5.2 What is 'shape' and how can we measure it?

It is widely accepted that shape is the result of processes (Hagget and Chorley, 1969; Goudie, 1990; McLellan and Endler, 1998) and thus, significant effort has been devoted to measuring, characterising and differentiating the shape of objects in many disciplines. There is a large body of literature devoted to the statistical analysis and classification of shape in fields such as:

- *Biology* e.g., measuring the variability in shape of animal skeletons (**Figure 5.1**) (Adams and Rohlf, 2000).
- *Computer Science* e.g., facial recognition from the distance between facial features (Slice, 2007).
- *Landscape Ecology* e.g., measuring the shape characteristics of different land cover types (Li *et al.,* 2001).
- *Remote Sensing* e.g., recognition and extraction of features (Benediktsson *et al.,* 2003).

Two-dimensional (2D) planimetric shape is all the geometrical information that describes an object that remains when location (translation), scale (dilation) and rotation are filtered out from an object (Kendall, 1984). Thus the first step in analysis of shape is to move from a complex, irregular polygon or set of pixels (i.e., a 'descriptive' shape) to a set of indices that can be quantified and computed, and most importantly, compared to other shapes by filtering out the effects of location, scale and rotation. This task is not trivial. Methods to digitise and classify shapes vary depending on the application, but include:

- *Geometric Morphometric Landmarks*: Digitising a set of points on a shape and looking at relationships between those points (e.g., relative distances and angles). Widely used in the field of biology for comparing the shape of organisms (e.g., **Figure 5.1**).
- *Shape indices*: Examining various relationships between area and perimeter of shapes. Widely used in the field of landscape ecology to describe the structural composition of landscapes (e.g., **Figure 5.2**).

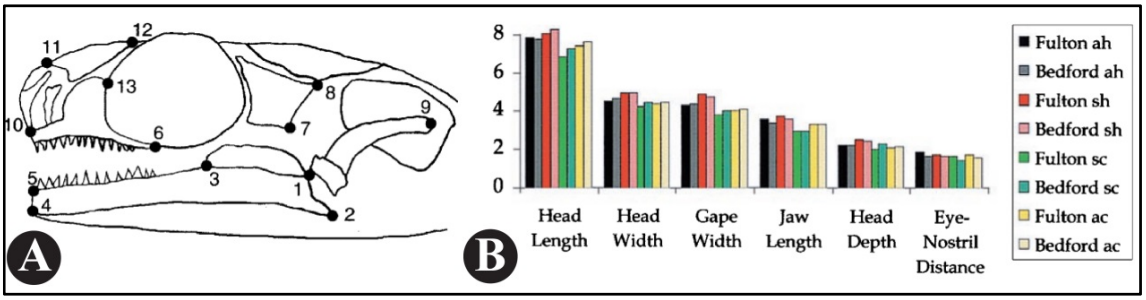


Figure 5.1 *Example of a geomorphic morphometric technique for the shape of salamander skulls (Figure from Adams and Rohlf, 2000).* ***(A)*** *The landmark points (points 1-12 in* ***A****) corresponding to different skull features are digitized; this is repeated for different salamanders from different locations and morphometric features such as the head length (distance between points 5 and 9) calculated.* ***(B)*** *The mean values of different morphometric features of salamanders are plotted on the y-axis and compared for different locations and populations of salamanders listed in the legend where location is first given, and then ah/sh/sc/ac correspond to different populations of salamander) to explain differences in skeletal development caused by availability of food at different locations.*

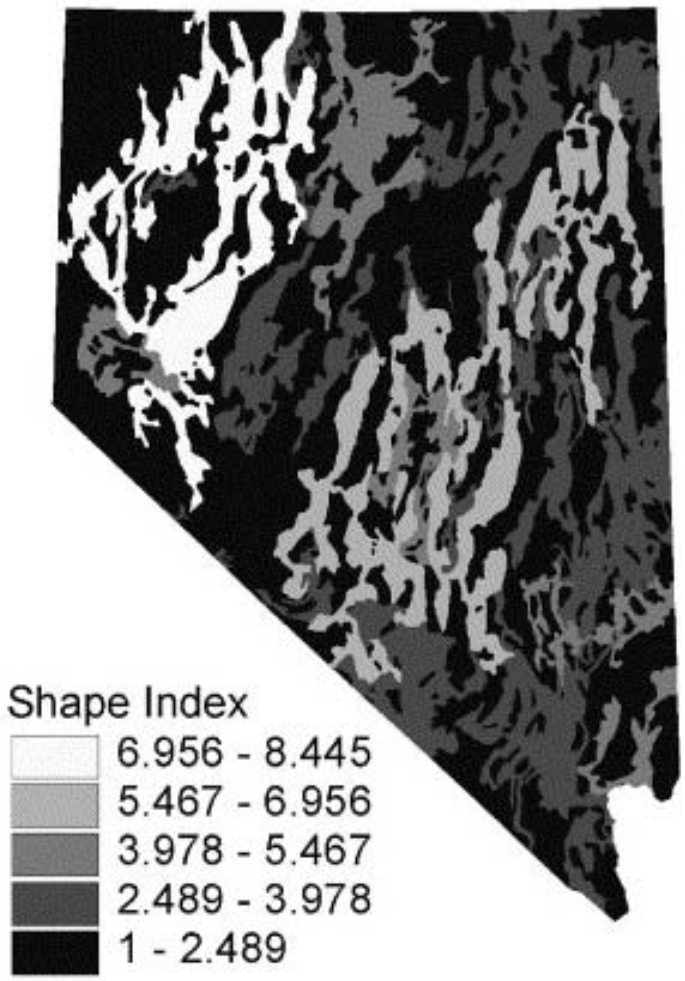


Figure 5.2 *Example of a shape index technique applied to geologic units (mapped at 1:2,500,000 scale) for the entire state of Nevada, USA (Figure from Raines, 2002). Mean shape index is calculated for each geologic unit as the sum of all geologic unit polygon perimeters divided by the square root of the total area for that geologic unit, and gives an indication of the complexity of a shape. A shape index value of 1.00 denotes a circular shape and values > 1.00 indicate more complex shapes. In this example, the alignment of zones with similar shape index values in a northeast-southwest direction correlates with the location of allochthonous terranes and potentially gives insight into processes of faulting and thrusting in this region.*

Although the geometric morphometrics approach may be considered more powerful for accurately quantifying the shape of organic forms, the process requires detailed digitization of a set of landmark points for each polygon. Whereas, shape indices are often easily calculable from simple measures such as area and perimeter. Forman (1995 p.135) notes that "no single measurement or index can unambiguously differentiate all shapes", but optimal methods will satisfy the following criteria:

1. *Be easy to calculate.*
2. *Be applicable to the entire region* of interest.
3. *Allow us to quantitatively differentiate different types* of shape.
4. *Allow us to plot a shape* based on the information given in that index.

In the following **Section 5.3** we review the literature discussing what shape landslides are, finding that a common shape index to describe landslide shape is length-to-width ratio (L/W), which also satisfies Forman's (1995) criteria listed above. In **Section 5.4** we introduce the landslide inventories used in analysis in this **Chapter 5**. In **Section 5.5** we experiment with ways to measure L/W of landslides in a method that can be systematically and rapidly applied to large inventories of landslides. In **Section 5.6** we investigate how the shape of landslides varies by looking at statistical patterns in L/W, particularly in relation to landslide area (A_L).

5.3 What shape are landslides?

Shape can refer to one-, two- and three-dimensional objects, and indeed, a landslide is a 3D object. Landslide shape can be measured in different ways:

- *3D analysis* of landslide volume. Examples include: Parise and Guzzi, (1992) and Klar *et al.* (2011).
- *2D analysis on the vertical plane* (e.g., relationship between landslide length and vertical depth or height). Examples include: Hovius *et al.* (1997), Stark and Guzzetti (2009) and Milledge *et al.* (2012).
- *2D analysis on the horizontal plane* (e.g., relationship between landslide length and width). Examples include: Barton *et al.* (1983), Parise and Jibson (2000) and Rickli and Graf (2009).

We focus here on the analysis of two-dimensional shape of landslides on the horizontal plane (i.e., the shape 'along' the plane that the landslide has formed), including both the source area and the run-out zone. This is because we later on aim to create 'synthetic'

triggered landslide event inventory maps with realistic landslide shapes and sizes in the landslide-road impact model (LRIM, **Chapter 7**). The landslide inventory maps we have available (which will be introduced in **Section 5.4**) are two-dimensional planar representations of real landslides, where both the source and run-out zone of individual landslides are mapped as one polygon. Thus, we focus our analysis on how to quantify and model the shape of landslides within the confines of the data we have available.

Although there have been considerable advances in our ability to quantify and classify shapes in the past few decades (Adams *et al.*, 2004; Slice, 2007), and there are agreed definitions of landslide dimensions (Flageollet, 1996; Hungr *et al.*, 2014) within the landslide literature, 'shape' is commonly described qualitatively, based on type of movement and material involved (e.g., Dikau *et al.*, 1996; Hungr *et al.*, 2014). Terms such as 'lobate', 'elliptical', 'isosceles triangle', 'long and thin', 'irregular', 'elongate', 'rectangular', 'spoon-like', 'amphitheatre' and 'cone-shaped' appeared frequently in the literature to describe landslide forms (from literature reviewed in **Table 5.1** and more broadly, the literature reviewed throughout this thesis). Indeed, geomorphologists would most likely be able to approximately identify different types of landslide shapes if provided with photographs or other representations of their form. **Figure 5.3** shows seven examples of common types of landslide shape identified in a 1998 Hurricane Mitch triggered landslide inventory demonstrating a variety of different forms and processes. **Figure 5.4** shows seven photographs of loosely-analogous landslides from the USGS archive of landslide photographs (note the landslides in **Figure 5.3** and **5.4** are not the same).

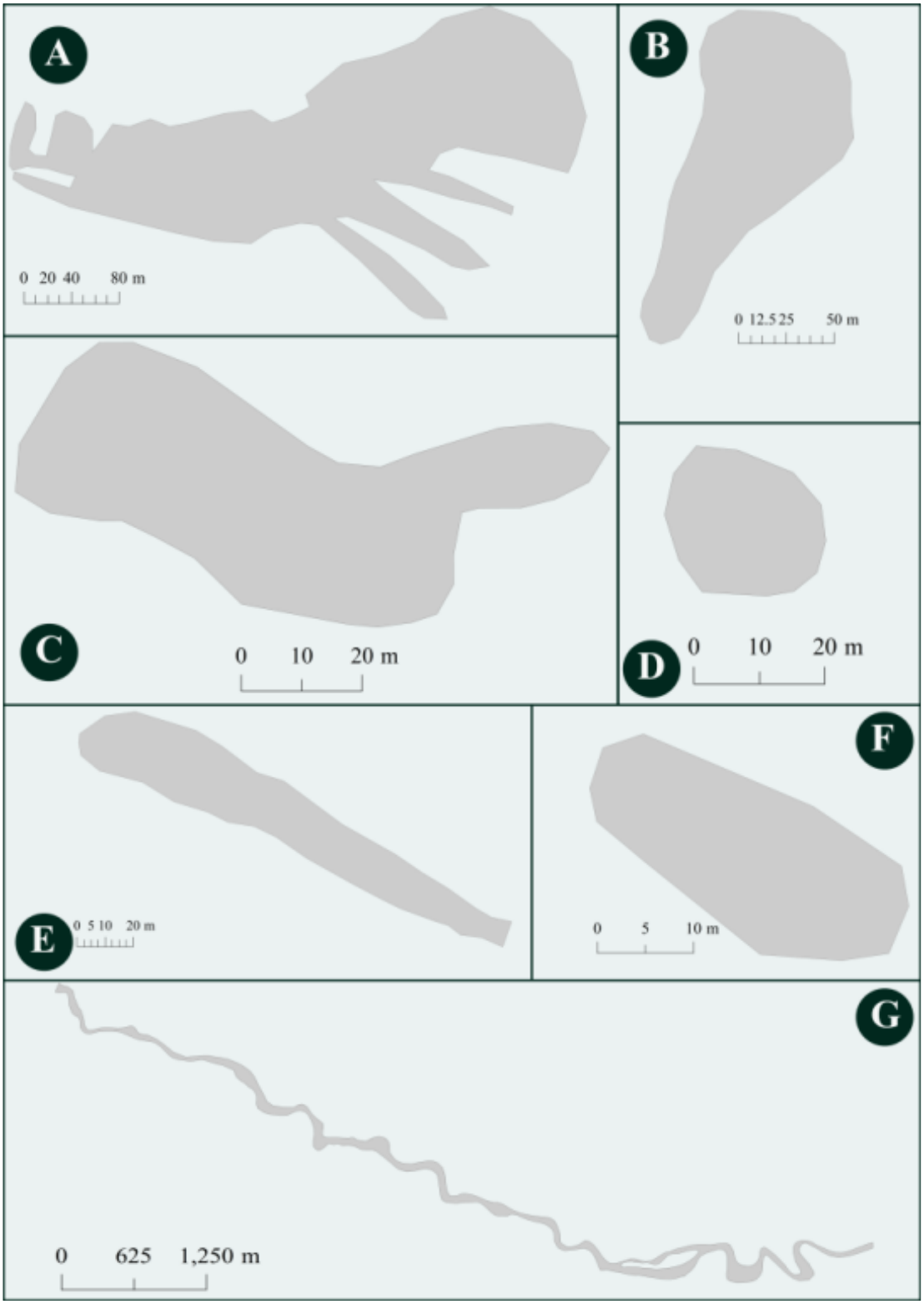


Figure 5.3 *Examples of seven commonly observed types of landslide shape and descriptions based on our visual analysis of a landslide inventory. The landslide polygons are taken from an inventory (Bucknam et al., 2001) of 9,594 landslides triggered by 1998 Hurricane Mitch in Guatemala.* ***(A)*** *A complex landslide with a very irregular, branched shape.* ***(B)*** *A roughly symmetrical 'wedge' shaped landslide that narrows in the flow direction.* ***(C)*** *A landslide that is initially approximately symmetrical and regularly shaped but appears to have an additional 'lobe' which we attribute to the landslide flowing around a topographical feature.* ***(D)*** *A compact, nearly circular landslide.* ***(E)*** *A relatively long, thin landslide with a fairly regular shape.* ***(F)*** *Similar to (E) but more compact and broadens in the direction of flow.* ***(G)*** *A very long, thin, sinuous shape typical of debris flows which follow channel morphology.*

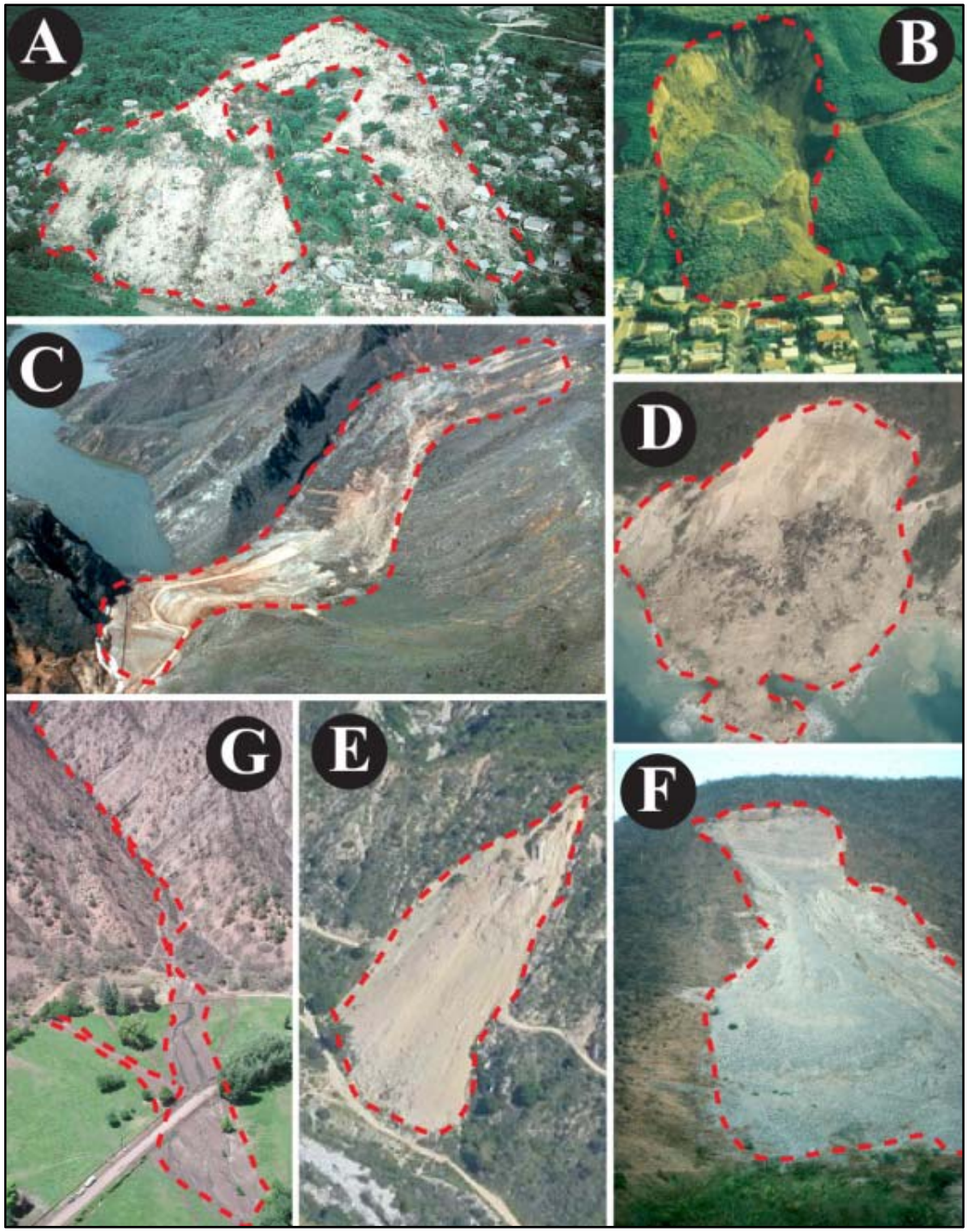


Figure 5.4 *Photographs of commonly observed types of landslide shape, which are loosely analogous to the landslide inventory polygons shown in* ***Figure 5.3****. Landslide information and Image Sources:* ***(A)*** *Photograph of the Mameyes, Puerto Rico rock slide, 1985 by R.W. Jibson, Courtesy of the U.S. Geological Survey****.*** ***(B)*** *Photograph of the La Conchita, California, 2005 by complex slump-earth flow by R.L. Schuster, Courtesy of the U.S. Geological Survey.* ***(C)*** *Photo of the Thistle Utah complex earthflow, 1983 by R.L. Schuster, Courtesy of the U.S. Geological Survey.* ***(D)*** *Photo of the South Coast Haiti landslide, 2010 by E.L. Harp, Courtesy of the U.S. Geological Survey.* ***(E)*** *Photo of a shallow landslide in the San Gabriel Mountains, California, 2005, Courtesy of the U.S. Geological Survey.* ***(F)*** *Photograph of a rockslide in Guerrero, Mexico, 2007 by Wikipedia User Adrignola,* ***(G)*** *Photograph of debris flow near Glenwood Springs, Colorado, 2002, by A. Holland-Sears, U.S.D.A, Courtesy of the U.S. Geological Survey. Those images that are 'courtesy of the U.S. Geological Survey' were all obtained via the USGS archive of landslide photographs (USGS, 2013).*

From **Figures 5.3** and **5.4**, we observe that landslide shape is typically 'organic': having curved edges and asymmetric borders. Shape varies depending on the interrelationship of local conditions (e.g., topography, geology, vegetation, antecedent precipitation, soil moisture) and physical processes, with both largely determining the type of landslide. For instance, individual planar or rotational slides and flows may be more symmetric and roughly elliptical (such as **Figure 5.4 B**, **D**, **E**, and **F**), although if they intersect with topographic features, such as **Figure 5.4C**, their path may be diverted and result in a less symmetric shape. Debris flows such as **Figure 5.4G** tend to follow channel morphology (Corominas *et al.*, 1996). More complex processes may result in very irregular shapes such as the 1985 Mameyes landslide in Puerto Rico (**Figure 5.4A**). This landslide was triggered by heavy rainfall and interactions with anthropic activity. In this case, the earth failed in three phases: (i) two translational slides, (ii) followed by a rock fall on the western part of the slide, which resulted in the rupture of human-made water pipes, (iii) the rupture of these water pipes in turn triggered flows (Jibson, 1992). It is the interaction and coalescence of these individual landslides, each controlled by different processes that have resulted in a complex, branched shape. Moreover, in this complex example (**Figure 5.4A**), the judgement of the geomorphologist creating the map also will affect the mapped shape of the landslide, with regards to:

- *Individual shapes*. It may be possible to delineate separate failures and map these individually.
- *Combined shapes*. The failures may be combined into one irregular shape, outlining where land has been disturbed (as shown by the red outline in **Figure 5.4A**).
- *Approximated shapes*. The mapmaker could draw a more regular line around the entire feature. This would result in an approximately triangular landslide shape encompassing most of the hillside, some of which did not actually fail but is surrounded by areas that did fail.

Landslide horizontal planar shape clearly varies quite considerably, both with process, but also potentially with mapping technique. We now discuss how landslide shape might be measured. We found that in the literature, landslide horizontal planar shape is typically measured from the ratio of landslide length-to-width, and this is referred to using various terms:

1. Length-to-width ratio
2. Length/Width
3. Aspect ratio

4. Elongation ratio
5. Geometrical characteristics

In **Table 5.1**, we show results of a literature search performed in Google Scholar using the terms listed above and summarise various ways in which landslide length-to-width ratio (L/W) is characterised. **Table 5.1** is broken into three parts: (**A**) articles that give some summary values of L/W but landslide shape is not the main focus of their paper; (**B**) articles that have performed a more in-depth analysis of landslide shape; (**C**) articles that state a relationship between landslide area and L/W.

Table 5.1 *Summary of existing literature discussing landslide length-to-width ratio (L/W).* ***(A)*** *Six papers stating values of L/W where this is not the main focus of the paper.* ***(B)*** *Seven papers investigating landslide shape in detail.* ***(C)*** *Six papers stating relationships between landslide L/W and landslide area.*

Authors	Location	Landslide Type	Inventory Type	Inventory Method	# Land-slides	Paper Theme	*L/W*	Notes	How Length is Measured	How Width is Measured
(A) Papers quoting values of landslide L/W but not main focus of the research in that paper										
Barlow *et al.* (2006)	Chilliwack River Basin, Vancouver, Canada	Debris slides; Rock slides; Debris flows	Historical	Aerial photo	60	Remote sensing; Inventory	Typically > 2.5		From crown to toe	Unknown
Dikau *et al.* (1996)	Various European locations	Single rotational slides	Unknown	Unknown	5	Description of landslide types	Min = 0.8 Median = 1.3 Mean =1.6 Max = 3.3		Maximum length perpendicular to slope	Maximum width parallel to slope direction
Gabet and Dunne (2002)	Santa Barbara, California, USA	Soil slips	Triggered Event: Rainfall	Aerial photo	31	Soil characteristics; Landslide volume	Min = 0.7 Median = 2.4 Mean = 2.5 Max = 5.6		Unknown	Unknown
Martel (2004)	US Continental Margin (submarine)	Submarine	Historical		11	Model; Landslide initiation; Submarine landslides	Min = 0.28 Median = 1.26 Mean = 1.47 Max = 3.90		From crown to toe	Perpendicular to length
Martha *et al.* (2010)	Mandakini River Catchment, Uttarakhand, India	All; Translational rock slides; Rotational rockslides; Debris flows; Shallow translational rock slides	Triggered Event: Rainfall	Satellite	466	Remote sensing; Inventory	Typically < 3		From crown to toe	Unknown
Moine *et al.* (2009)	Barcelonnette basin, Alpes-de-Haute-Provence, France	All; Translational slides; Rotational slides; Earthflows; Block slides	Historical	Aerial photo	156	Remote sensing; Inventory	Min = 0.5 Max = 5		From crown to toe	Unknown

Authors	Location	Landslide Type	Inventory Type	Inventory Method	# Land-slides	Paper Theme	*L/W*	Notes	How Length is Measured	How Width is Measured
(B) Papers investigating landslide morphometry in detail										
Barton *et al.* (1983)	Christchurch Bay Coastal Cliffs, United Kingdom	Cliff top slumps	Unknown	Unknown	42	Morphometry of slumps and spalls	Min = 0.44 Median = 4.2 Mean = 5.2 Max = 15.2	Clear difference in dimensions between geological zones. Non normal distribution.	Effectively the width of the crown in plan view	Effectively the 'depth' of the crown in plan view
Parise and Jibson (2000)	Santa Susana Quadrangle, Northridge, California, USA	All	Triggered Event: Earthquake	Aerial Photo	1562	Morphometry; Susceptibility	Mean (Individual slides) = 2.6 Mean (Complexes) = 1.2		From crown to toe	Perpendicular to length, maximum extent
Pourgha-semi *et al.* (2013)	North of Tehran, Iran	All	Historical	Aerial Photo; Satellite ; Field Survey	528	Morphometry; Fractals	Min = 1.001 Max = 6.084		Minimum distance from toe to crown in the downslope direction	Maximum breadth perpendicular to length
Quinn *et al.* (2011)	Saint Lawrence Lowlands, East North America	Large landslides in sensitive clay	Historical	Archive; Aerial Photo; Satellite	~62	Morphometry; Impact	Min = ~ 0.1 Max= ~10	Min (for spreads) = ~ 0.1 Max (for spreads) = ~1.4 Min (for flows) = ~ 0.7 Max (for flows) = ~10	Unknown	Width at outlet
Rickli and Graf (2009)	Central and East Switzerland	Natural; Shallow	6 Triggered Events (Rainfall)	Aerial Photo	522	Morphometry; Land use	Min = 0.95 Median = 1.44 Mean = 1.36 Max = 2.05		Maximum extent from crown to toe, No run out	Maximum extent perpendicular to length, No run out

Authors	Location	Landslide Type	Inventory Type	Inventory Method	# Land-slides	Paper Theme	*L/W*	Notes	How Length is Measured	How Width is Measured
Süzen (2002)	Asarsuyu, Turkey	All	Multi Temporal (4 periods)	Aerial Photo	154	Morphometry, Temporal change	Min = 1.47 Mean = 1.47 Max = 3.85		Longest Axis	Unknown
Yang and Lee (2006)	Central Taiwan	All	2 Triggered Events (Rainfall and Earthquake)	Aerial Photo	468 (Rainfall) 189 (Earthquake)	Morphometry: Fractal dimension	Rainfall triggered < 8 Earthquake triggered < 5		Unknown	Unknown
(C) Papers giving deterministic relationships between landslide area and dimensions										
Casadei *et al.* (2003)	Coos Bay, Oregon, USA	Shallow landslides	Historical	Unknown	90	Deterministic model of landslide size		$L = (B - 2/W)^{-1}$ $W = 2(B(1/L))^{-1}$ *where B is a function of cohesion, soil density and slope* *L* and *W* functions of soil and slope. Minimum width for landslide to fail is a function of root cohesion.	Unknown	Unknown
Frattini and Crosta (2013)	Trento, Italy	All	Multi Temporal	Aerial Photo; Lidar	4,175	Physical reasons for observed frequency size statistics		$W = L/2$	Unknown	Unknown
Guthrie *et al.* (2008)	British Columbia, Canada	Debris flows and slides	N/A	N/A	N/A	Cellular automata model of landslide size	*W* typically = 5 – 50 m	$L = 0.76A_L^{0.66}$ Assuming that landslides initiate mid-upper slope.	N/A	N/A

Authors	Location	Landslide Type	Inventory Type	Inventory Method	# Land-slides	Paper Theme	*L/W*	Notes	How Length is Measured	How Width is Measured
Hovius *et al.* (1997)	Western Southern Alps, New Zealand	Falls, slumps, Slides and Debris flows	Multi Temporal	Aerial Photo	4,984	Calculation of landslide volume and denudation	Typically = 2	$W = A_L^{0.5}$	Unknown	Unknown
Milledge *et al.* (2014)	Multiple: Japan, United Kingdom, USA: Oregon, California, Appalachian Mountains	All	Mixed: Multi-temporal, Seasonal and Event	Mixed	Unknown	Deterministic model of landslide size and shape	Min = 0.6 Max = 10.0 Generally < 3.0	*L/W* a function of soil depth, cohesion and friction	Unknown	Unknown
Reneau and Dietrich (1987)	Central California Coast, USA	Shallow debris flows	Triggered Event: Rainfall	Unknown	61	Deterministic model of landslide size	Min = 0.6 Mean = 1.8 Max = 4.2	$L = \frac{2\,S_r k\,W}{W - 2\,S_r k}$ Where S_r = Root strength, and k is a function of soil and slope	Unknown	Unknown

From the literature summarised in **Table 5.1**, we note the following trends:

- *Of all observed values summarised in* ***Table 5.1****, L/W ranges 0.28 – ~10* (excluding Barton *et al.* (1983) where *L/W* is defined differently), and the mean of the mean values stated = 1.7 (Standard Deviation (*SD*) = 0.5).
- *It is not always stated how L and W are defined* (e.g., is *L* the longest axis or the length in the downslope direction).
- *Many of the analyses performed are on relatively small sample sizes* (i.e., $N_{LS} < 200$ landslides).
- *The analyses rarely consider the 'population' of landslide shapes*, more frequently focusing on a subset of landslide types or using historical inventories (where small area landslides are known to be proportionally under sampled (discussed in **Chapter 2**).
- *A small number of authors approximate landslide shape scaling with area.*
- *Some authors have shown that landslide L/W differs with type of landslide.*
- *Some authors have proposed that landslide shape is a product of soil parameters and slope.*

The examples of landslide shape listed in this section and **Table 5.1** are generally based on field measurements of relatively small populations of landslides. As observed with the general behaviour of landslide frequency-size statistics (**Chapter 4**), we would like to synoptically test whether there is some 'general' statistical behaviour of landslide length-to-width ratio that can be used to model the shape of landslides in the landslide-road impact model (LRIM, which we develop in **Chapter 7**); and perhaps more widely used for probabilistic forecasting of landslide shape across many regions. Thus, in the following **Section 5.4**, we introduce seven triggered landslide event inventories from different regions across the world with different triggering mechanisms, containing a total of 23,176 landslides which we use for analysis of length-to-width ratio in this chapter.

5.4 Landslide Inventories Used

To observe the full variety and numbers of landslide shapes, it is important that we work with relatively complete inventories, where the majority of landslides have been mapped across all scales (Malamud *et al.,* 2004a). Because of issues with inventory production methods described in **Chapter 2**, it is a common issue that in many inventories, the smallest landslides are proportionally under sampled (Guzzetti *et al.,* 2012). Because of these known issues with landslide inventories, we initially adopt relatively strict criteria for initial inventory selection (based on Harp *et al.* (2011)'s criteria outlined in **Table 2.5**, **Chapter 2**) to ensure a statistically representative sample. The criteria are:

(i) *Inventory should be produced from medium to large scale aerial photography* (< 1:60,000) combined with field investigations.

(ii) *Reconnaissance and/or imagery should cover the majority of the extent* of the affected region.

(iii) *There should be a relatively short time lag* between the occurrence of the triggering event and the imagery/field mapping (< 6 months).

(iv) *The distribution of individual landslide areas must follow reasonably well already established landslide frequency size distributions* (e.g., Stark and Hovius, 2001; Malamud *et al.,* 2004a) to ensure representative sampling of all landslide sizes.

In addition to the above four criteria, we also required inventories with large numbers of individual landslides (N_{LS}). This is because in **Section 5.6.3**, we investigate the probability density distribution of landslide length-to-width ratio within categories of landslide area (A_L) (e.g., $10 \leq A_L < 100$ m^2). For robust statistical analysis of the probability density distribution of L/W within each A_L category, we required a reasonable number of observations within each A_L category.

A common heuristic for the required sample size to fit a statistical distribution is 10–20 observations per parameter being fit to the data (Harrell, 2002). From initial trials, a three-parameter inverse Gamma distribution appeared to be an appropriate data model for landslide length-to-width ratio probability density distribution, and thus we require at least 30–60 observations in each landslide area category. We will find later (**Section 5.6.3**) that assuming landslide area follows an inverse Gamma probability distribution, and ten landslide area categories (with some area categories having more landslide observations than others), requires an overall number of landslides in the inventory of 7,000–14,000 observations.

Of available triggered event inventories, the following meet our criteria:

(i) *Northridge:* 11,111 landslides triggered by the (M_W = 6.7) 1994 *Northridge* Earthquake in California, USA (Harp and Jibson, 1995). Landslides were mapped immediately after the earthquake using field surveys and 1:60,000 scale aerial photography. A subset of the inventory is shown in **Fig 5.5A**.

(ii) *Guatemala:* 9,594 landslides triggered by Hurricane Mitch in *Guatemala* in 1998 (Bucknam *et al.*, 2001). Landslides were mapped using 1:40,000 scale aerial photography taken in 2000. A subset of the inventory is shown in **Fig 5.5B**.

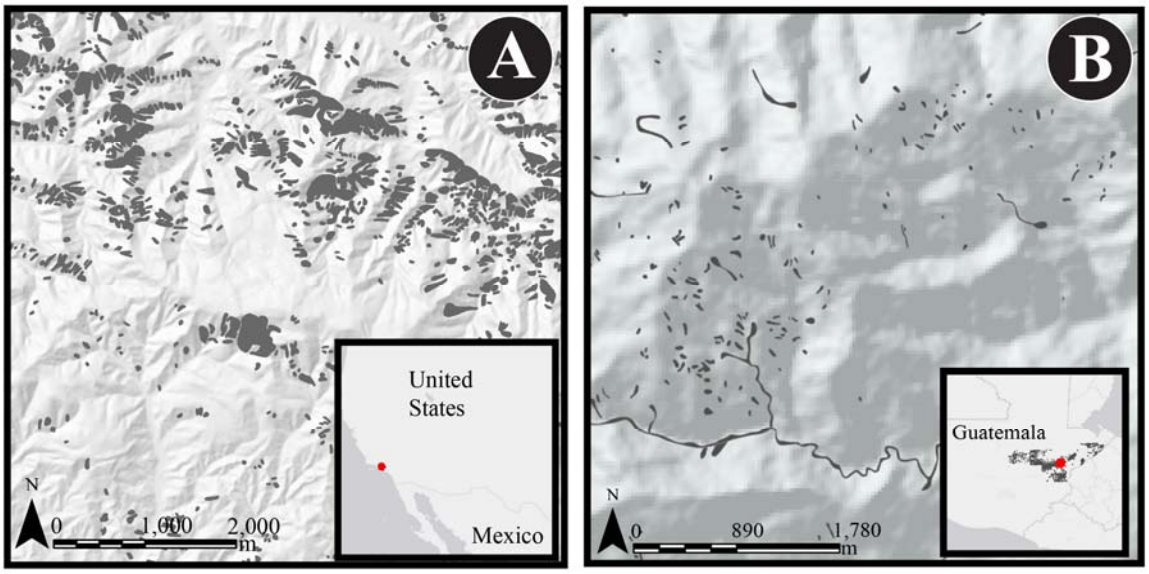


Figure 5.5 *Subsets of triggered landslide event inventories.* ***(A)*** *Subset of Northridge landslide event inventory where 11,111 landslides were triggered by the 1994 Northridge Earthquke in California, USA (Harp and Jibson, 1995).* ***(B)*** *Subset of Guatemala landslide event inventory where 9,594 landslides were triggered by Hurricane Mitch in Guatemala (Bucknam et al., 2001). Digital elevation models from USGS (2006). Inset maps (bottom right) show the extent of the entire inventory (grey polgons) and the location of the zoomed inventory subsets within each country.*

We call the Northridge and Guatemala inventories (**Figure 5.5**) 'high confidence' as they are substantially complete and contain large numbers of individual landslides, allowing robust statistical analysis when split into landslide area categories. These inventories are used to experiment with different methods of approximating landslide shapes as ellipses (**Section 5.5**) and exploring patterns in landslide frequency size statistics (**Sections 5.6.1 – 5.6.3**). After experimenting with these two inventories and establishing a workflow for analysis, in **Section 5.6.4** we then apply the analysis to five other triggered event inventories (**Table 5.2**) which do not fully meet the criteria listed above, either because of issues of sample size or inventory production method. This is done to test whether there are some general trends in landslide length-to-width ratio observed in various triggered landslide events. In **Table 5.2**, we categorise landslide inventories by our 'confidence' in sample size and production methods (which is green for high confidence; amber for medium confidence; and red for low confidence) and also include some notes about the predominant types of landslides observed in each inventory. The potential link between landslide type, processes and shape is discussed in **Section 5.7**.

Table 5.2 *Summary of triggered landslide inventories analysed in* ***Chapter 5****. Landslide inventories are categorised by 'confidence', based on our understanding of the inventory production methods, amount of documentation about the inventory, fit to already established landslide frequency size statistics and sample size. "# Landslides accepted" refers to the method to calculate ellipticity goodness-of-fit (e) described in* ***Section 5.5.2*** *and denotes what proportion of the inventory landslide polygons are well modelled by an ellipse.*

Inventory Name	Reference	Total # landslides	# landslides accepted (Sect. 5.5.2)	Trigger	Notes on Type of Landslides	Notes on 'confidence' in inventory
Collazzone (Italy)	Cardinali *et al.* (2000)	413	397 (96 %)	Rapid Snowmelt1997	**Broadly low mobility** 62% small shallow slips. Some larger earth flows and deep seated slides. Occurring on slopes > 20° with shear surface parallel to slope. Clay soil prevented debris flows.	**Medium confidence:** Small size dataset. Considered significantly 'complete' (e.g., Malamud *et al.*, 2004a).
El Salvador	Crone *et al.* (2001)	621	514 (83 %)	Hurricane Mitch, 1998	**Significant proportion high mobility landslides** Almost entirely shallow slides in unconsolidated material. Generally travelled from top to base of slope. Many transitioned into debris flows.	**Low confidence:** Medium size dataset. Imagery covered > 50 % of El Salvador. Authors acknowledge the inventory is not entirely complete. Contains some landslides that occurred before Hurricane Mitch.
Guatemala	Bucknam *et al.* (2001)	9,594	8,031 (84 %)	Hurricane Mitch, 1998	**Significant proportion high mobility landslides** Many small shallow translational slides, many mobilised into debris flows. Some large translational slides which transitioned into debris flows following channels. Occurring on moderate to steep hillslopes across diverse geology.	**High confidence:** Large size dataset. Considered to be 95 % complete (Bucknam *et al.* 2001).
Liguria (Italy)	Mondini (2014)	537	357 (66 %)	Rainfall (2011)	**Significant proportion high mobility landslides** All shallow landslides. Translational slides, earth flows, soil slips and rotational slides. Slides located at the foot of slopes. Flows channelled in secondary streams.	**Low confidence:** Medium size dataset. Created from semi-automatic interpretation of satellite imagery. Some small landslides missed (~8.7 % compared to photo interpretation).

Inventory Name	Refer-ence	Total # land-slides	# land-slides accept-ed (Sect. 5.5.2)	Trigger	Notes on Type of Landslides	Notes on 'confidence' in inventory
Northridge (USA)	Harp and Jibson (1995)	11,111	9,441 (85 %)	North-ridge Earth-quake, 1994	**Generally low mobility landslides** Many shallow, highly disrupted falls/slides. Occurred in weakly cemented sediments. Most larger slides. travelled < 50 m, some travelled up to 200 m.	**High Confidence:** Large size dataset. Considered significantly complete (e.g., Malamud *et al.*, 2004a).
Rocca-fluvione (Italy)	LAMPRE (*personal communi-cation,* 2014)	(960*) 851	750 (88 %)	Rainfall and snowmelt 2013	**Visual analysis suggests generally low mobility landslides** Not yet documented.	**Medium Confidence:** Medium size dataset. Not yet fully documented. Created from detailed field inventory. Possible bias towards landslides visible from road. Frequency-area rollover occurs at very small landslide area (100 m^2).
(Shu-Wa) Taiwan	Chen (2014, personal comm-unication)	149	140 (94 %)	Typhoon Megi, 2010	**Generally high mobility landslides** Rockfalls, debris slides, rock slides and debris flows. Occurring in colluvium in upper slope portions.	**Low confidence:** Small size dataset. Not well documented. Matches well established landslide frequency-size statistics.

*(*109 landslides removed due to issues with topology)*

5.5 Calculating the Length-to-Width Ratio of Landslides

The two large triggered event inventories we use for initial analysis (Northridge and Guatemala) contain a large number of landslides (N_{LS} > 20,000 individual polygons in total) so it is not practical to measure each polygon shape individually. Instead, we develop a method that can rapidly and systematically be applied to all polygons in an inventory in an automated way. From the literature search and visual assessment of the inventories, we start with the hypothesis that many or most landslide polygon shapes can be abstracted to an ellipse (Hovius *et al.*, 1997; Martel, 2004; Marchesini, *et al.* 2009; Martha *et al.,* 2010, Pourghasemi *et al.,* 2013) and that variability in landslide shape can be approximated by variability in ellipse length-to-width ratio (L/W). Length (L) is the long axis of the ellipse, regardless of landslide travel direction and width (W) is the axis perpendicular to this, measured halfway along the long axis, shown in **Figure 5.6**. Although in reality it is possible to observe landslides where L/W < 1.0 (**Table 5.1**) (i.e., L in the travel direction is

shorter than perpendicular *W*) others have noted that this is generally a minor proportion of landslides in inventories (Gabet and Dunne, 2002; Rickli *et al.*, 2008; Marchesini *et al.*, 2009; Milledge *et al.*, 2014) and thus we assume this is a reasonable simplification for the sake of reducing processing time.

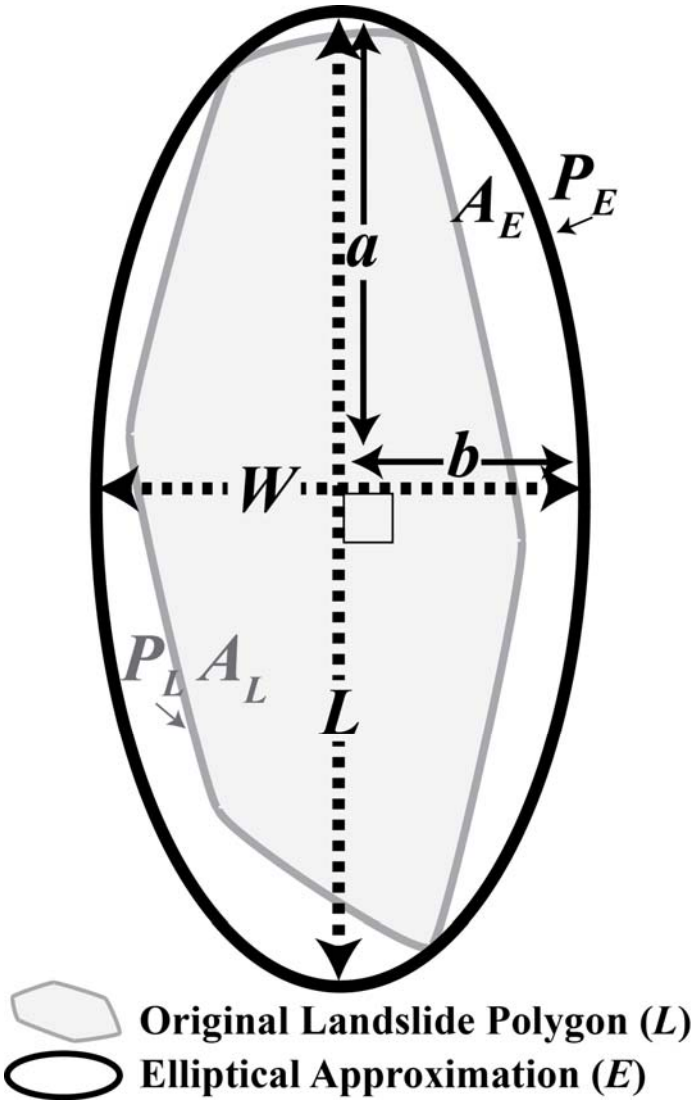


Figure 5.7 *Landslide and Ellipse shape parameters. In this example, an ellipse (black thick line) with area (A_E) and Perimeter (P_E) has been fit to a real landslide (grey polygon) with area (A_L) and Perimeter (P_L). The ellipse has a semi major axis (a) and semi minor axis (b). Ellipse length is equal to 2a and ellipse width equal to 2b. The ellipse length-to-width ratio is $L/W = a/b$.*

In **Figure 5.7** we show experimental ellipse approximations of landslide polygons and their resulting length-to-width ratio, compared to length-to-width ratios of the real landslide polygons measured by eye (which we acknowledge could have been measured in a variety of ways). **Figure 5.7** shows that L/W measured by eye and from elliptical approximations is roughly similar, and thus this is an appropriate way of measuring landslide shape.

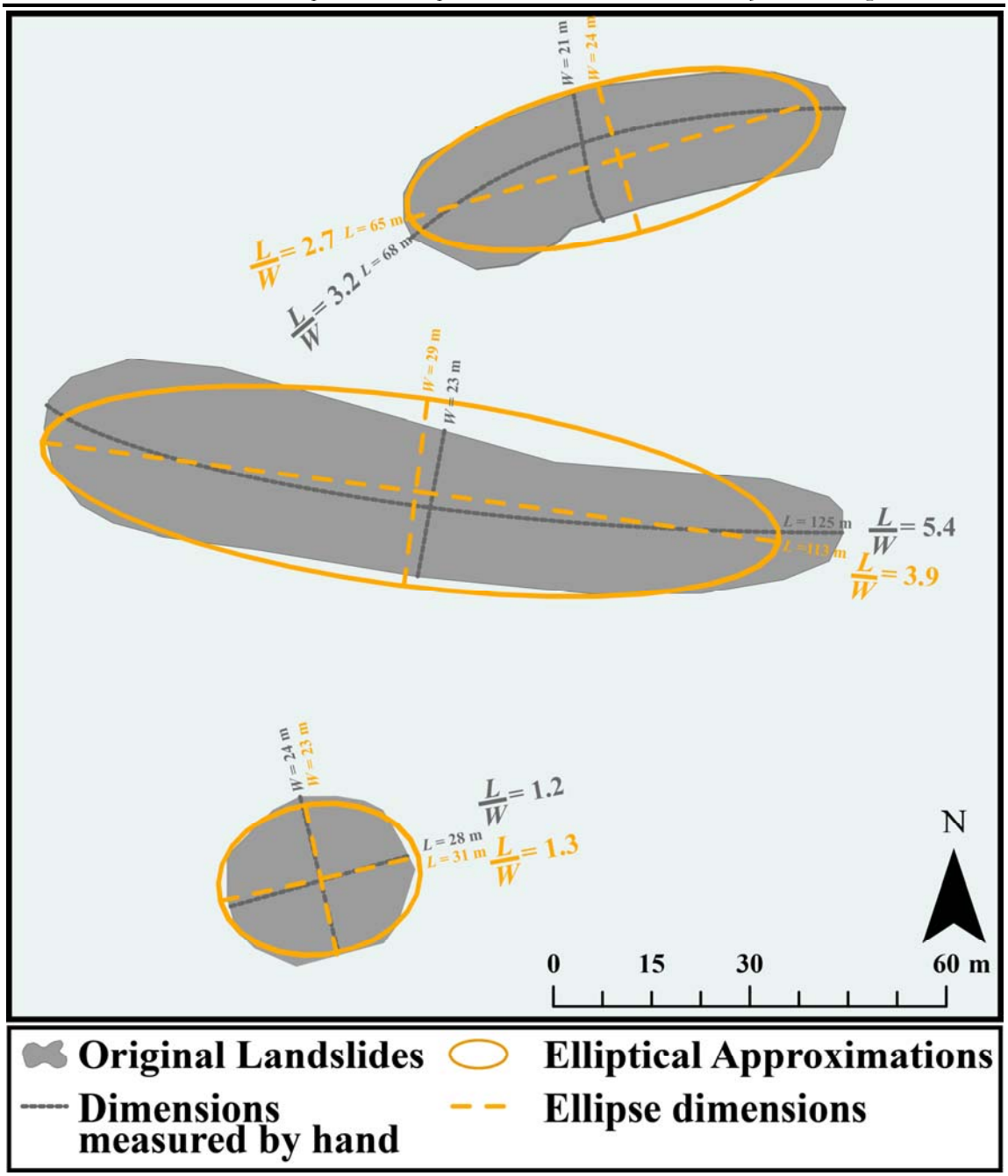


Figure 5.8 *Examples of elliptical approximations of real landslide polygons and the length-to-width ratio of those ellipses compared to one measure of length-to-width ratio of the true landslide polygon, done by eye. All three landslide polygons are from the Guatemala Hurricane Mitch inventory (Bucknam et al., 2001).*

In the following sections we establish a workflow to:

5.5.1 *Approximate an ellipse* from landslide polygons in a systematic way.

5.5.2 *Test the goodness-of-fit* of these elliptical approximations.

5.5.3 *Calculate the length-to-width ratio* of landslide elliptical approximations.

At the end of **Section 5.5** we have a robust methodology to calculate landslide ellipse L/W for landslide inventories. We then analyse the statistical distribution of L/W in **Section 5.6**.

5.5.1 Methods to Approximate Landslide Polygons to Ellipses

There are several ways in which an ellipse can be approximated from a polygon, and each will result in different values of L/W. We experimented with four methods to fit an ellipse to a shape, listed and described in **Table 5.3**. We applied Methods 1 – 3 (described in **Table 5.3**) to all landslides in both the Northridge and Guatemala inventories. We then tested the goodness-of-fit of each ellipse using the ellipticity (e) index (Lombardo, 2014), described in the following **Section 5.5.2**. We found that Method 4 (**Table 5.3**) was extremely computationally demanding and did not produce significantly improved elliptical approximations of landslides in terms of goodness-of-fit when applied to a small sample of landslide polygons, so Method 4 was not investigated further.

Table 5.3 *Different methods used in this chapter for approximating an ellipse shape from real landslide polygons. The notation Ef( ) is used to indicate 'ellipse fitting' method, with variables in brackets.*

Method	Notation	Description	Figure
1. Quadratic Ellipse of Landslide	E$f(A_L, P_L)$	Uses a quadratic equation (**Eq. 5.3**) to calculate ellipse L/W from perimeter (P_L) and area (A_L) of the landslide.	**5.9A**
2. Quadratic Ellipse of Convex Hull	E$f(A_{CH}, P_{CH})$	Uses a quadratic equation (**Eq. 5.3**) to calculate ellipse L/W from perimeter (P_{CH}) and area (A_{CH}) of a convex hull fit to the landslide. A convex hull is the minimum bounding area that encloses a polygon where all internal angles connecting vertices are convex (de Berg *et al.*, 2008).	**5.9B**
3. Quadratic Ellipse of Convex Hull scaled to A_L	E$f(A_{CH}, P_{CH})\rightarrow(A_L)$	Same method as E$f((A_{CH}, P_{CH})$, but the area of the ellipse is scaled to match the area of the original landslide (A_L)	**5.9C**
4. Standard Deviation Ellipse	E_{SD}	Performed within a GIS. Calculation of mean centre point of shape. X and Y axes are defined as one standard deviation of coordinates in the North and East directions (Yuill, 1971).	**5.9D**

The length-to-width ratio (L_E/W_E) for an idealised ellipse can be approximated from area (A) and perimeter of a shape (P). This is done using a quadratic equation, based on substituting the equations for ellipse area (**Eq. 5.1**) and an approximation of ellipse perimeter (**Eq.5.2**) in to one another (**Eq. 5.3**). This quadratic equation method was used by Malamud *et al.* (2004a) to calculate L/W of landslide polygons based on landside area

(A_L) and perimeter (P_L). The area of an ellipse (A_E) is calculated from the long (a) and short (b) axis of the ellipse (Heath, 1931) as:

$$A_E = \pi ab \quad (5.1)$$

Where: A_E = area of ellipse
a = long axis of ellipse
b = short axis of ellipse

Calculating the exact perimeter of an ellipse is complex, so it is common to use a variety of ellipse approximations based on the long (a) and short (b) axis of the ellipse. Here, we use Euler's (1773) formula (described in Michon, 2015) (**Eq. 5.2**), due to its relative simplicity, although note that this can be up to 10 % inaccurate for very long, thin ellipses:

$$P_E \approx 2\pi\sqrt{\tfrac{1}{2}(a^2+b^2)} = \pi\sqrt{2(a^2+b^2)} \quad \textbf{(5.2)}$$

Where: P_E = approximate perimeter of ellipse
a = long axis of ellipse
b = short axis of ellipse

By substituting **Eq. 5.1** into **Eq. 5.2** and vice versa using a quadratic formula, the length-to-width ratio of an ellipse (L_E/W_E) can be calculated from the area (A) and perimeter (P) of any shape (Malamud, *personal communication*, 2012):

$$\frac{L_E}{W_E} = \frac{4\pi A}{P^2 - \sqrt{P^2 - 16\pi^2 A^2}} \quad \textbf{(5.3)}$$

Where:
A = Area of a given shape.
P = Perimeter of a given shape.
L_E = Length of ellipse
W_E = Width of ellipse

In **Eq. 5.3**, the given shape could be the original landslide (A_L and P_L) or a convex hull fit to the landslide (A_{CH}, P_{CH}). The orientation and centre point from which each ellipse is mapped is the same in all cases. This is done using the minimum bounding geometry bounding box tool in ArcMap 10.1. For each landslide polygon, the minimum bounding rectangle that fully encloses the polygon is calculated, and the orientation of the landslide is aligned with the longest axis of this rectangle. The centre point from which the ellipse is plotted is calculated from the landslide polygon mean centroid, based on a weighted average of the polygon's Northing and Easting coordinates (N, E). In a small number of cases, this may be outside the original polygon extent.

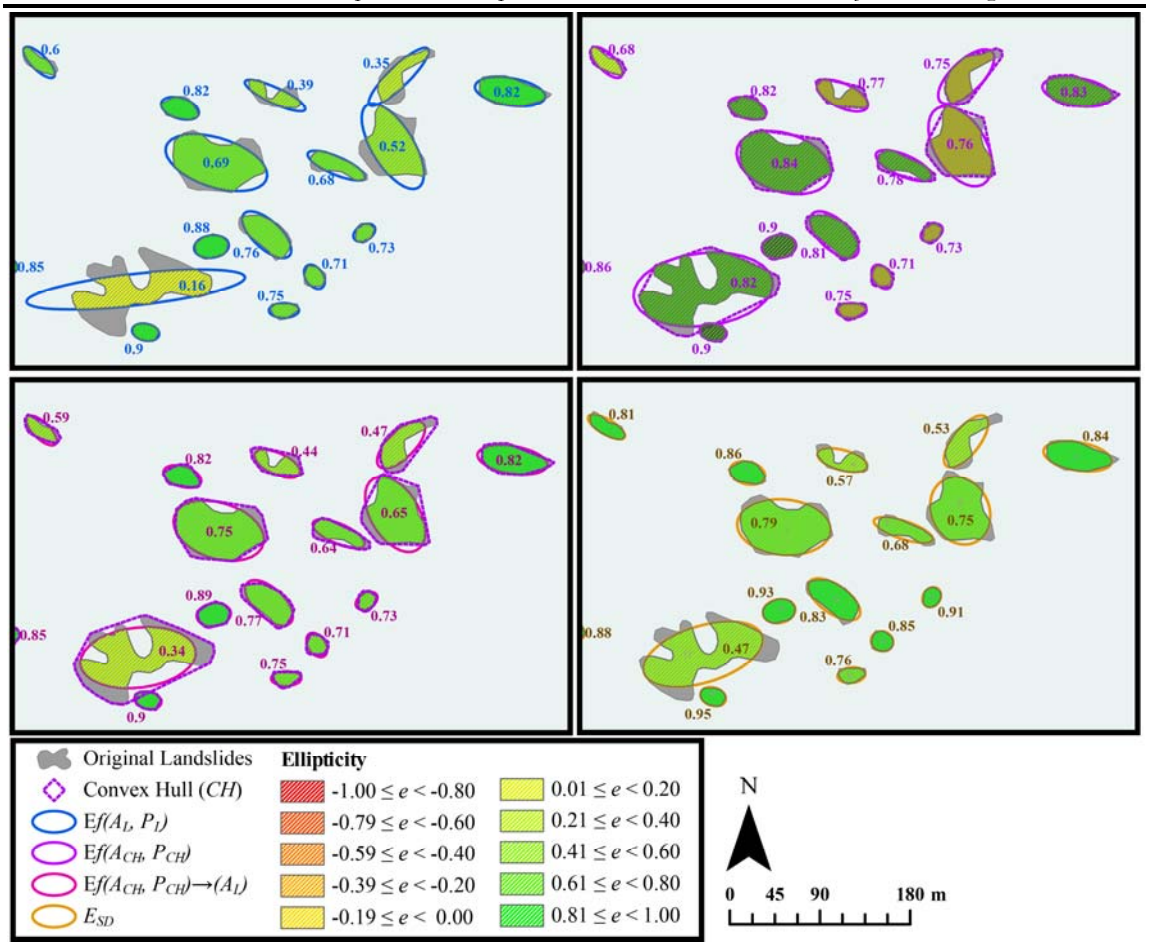


Figure 5.9 *Elliptical approximations of landslide polygons from the Guatemala Inventory and goodness-of-fit (ellipticity, e,* ***Eq. 5.2****). Each method of approximating an ellipse is described in* ***Table 5.2****.* ***(A)*** *Ellipse calculated from landslide area (A_L) and landslide perimeter (P_L).* ***(B)*** *Ellipse calculated from convex hull (CH) area (A_{CH}) and perimeter (P_{CH}).* ***(C)*** *As in (B) but ellipse area (A_E) scaled to match area of original landslide polygon (A_L).* ***(D)*** *Standard deviation ellipse calculated in a GIS.*

5.5.2 Testing the Goodness-of-fit of Elliptical Approximations of Landslide Polygons

As shown in **Figure 5.9**, many landslides are approximately elliptical in shape, but a small number are not. Landslides that are not well modelled by an ellipse tend to be mainly landslide complexes and debris flows, where complexes are defined as multiple landslides mapped as one polygon due to difficulty delineating the boundaries of individual landslides. To test the goodness-of-fit of ellipses fit to landslides using Methods 1-3 (**Table 5.3**), we use an ellipticity index *e* developed by Lombardo (2014), based on the area of intersection (A_I) between the original landslide area (A_L) and the elliptical approximation of the landslide where:

$$e = 1 - \frac{2(A_L - A_I)}{A_L} \quad \textbf{(5.4)}$$

e = Goodness-of-fit measure
A_L = area of original landslide polygon
*A_I = area of intersection (hashed areas in **Figure 5.9**) between original landslide polygon and idealised ellipse*

Ellipticity (e) ranges between −1.0 and 1.0. A value of e = −1.0 is a completely imperfect fit where there is no intersection (A_I = 0) between the original landslide polygon and the idealised ellipse. A value of e = 1.0 is a perfect fit (both the original landslide polygon and the idealised ellipse are exactly the same shape and align perfectly, so $A_I = A_L$). A small number of landslides will each have multiple values of e where the area of intersection between the original landslide polygon and elliptical approximation (A_I) is split into non-contiguous sections in the GIS. This generally reflects long, thin, sinuous landslides (i.e., debris or mud flows that follow river channel morphology) where an ellipse is a poor approximation of landslide shape. In these cases, e is set to null. Examples of ellipticity ranging from e = −1.00 to e = 0.99 are shown in **Figure 5.10** for Method 3 (ellipse approximated from the area and perimeter of a convex hull, scaled to the area of the original landslide) ($Ef(A_{CH}, P_{CH}) \rightarrow (A_L)$).

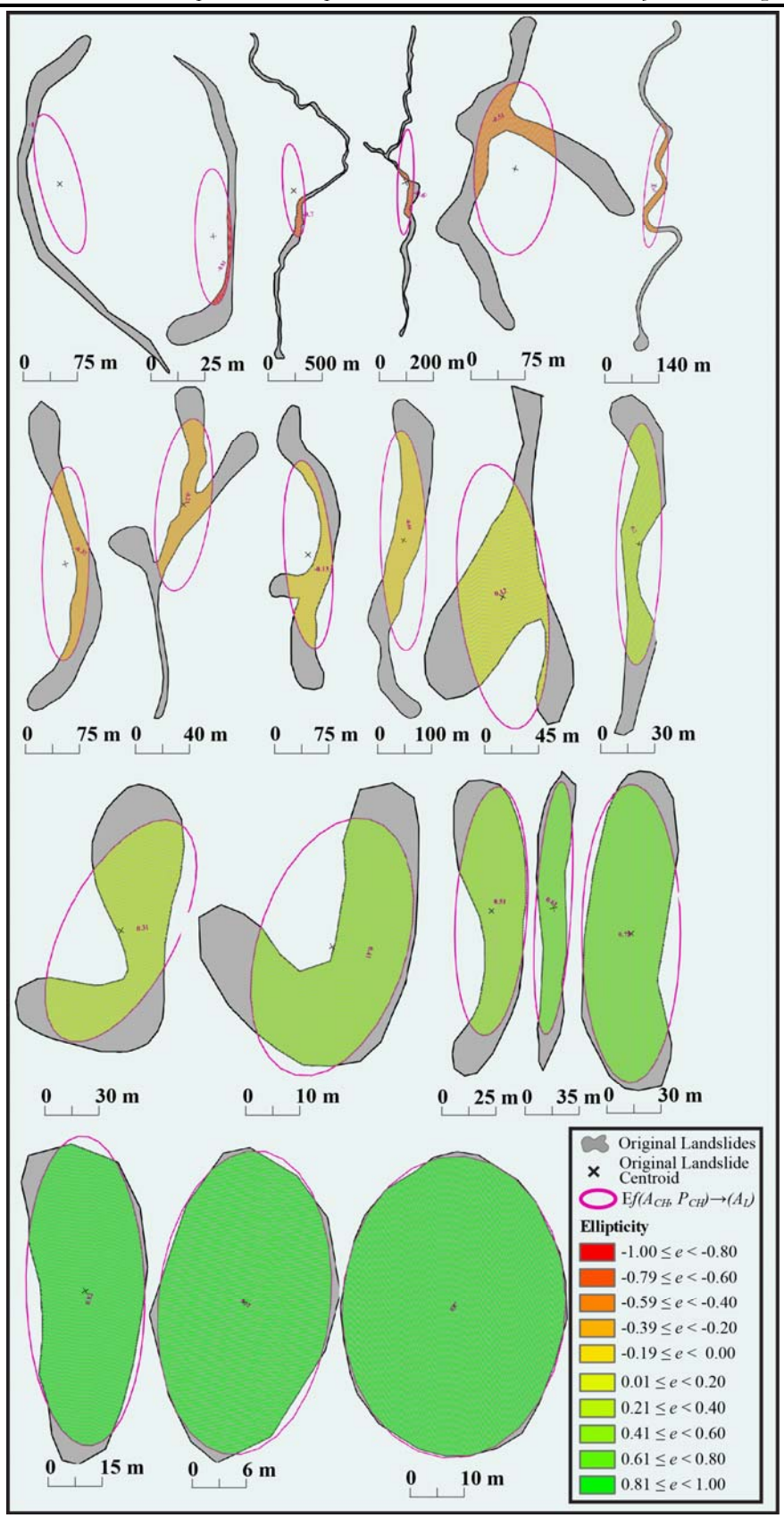


Figure 5.10 *Examples of elliptical approximations ($Ef(A_{CH}, P_{CH})\rightarrow(A_L, P_{CH})$) of real landslides with varying levels of goodness-of-fit (measured by ellipticity (e),* ***Eq. 5.4****). Examples of ellipticity are shown in e = 0.1 increments. Real landslide polygons are from the inventory of 9,594 landslides triggered by Hurricane Mitch in Guatemala (Bucknam et al., 2001).*

Shown in **Figure 5.11**, the probability density distributions of ellipticity values were compared for Methods 1 to 3. We observe that for both the Northridge **(A)** and Guatemala **(B)** inventories, the majority of landslide-ellipses have ellipticity values $e > 0.0$. For all three Methods 1–3, ellipticity peaks between $0.6 \leq e < 0.8$. For both locations, the method with the highest values of ellipticity is Method 2 ($Ef(A_{CH}, P_{CH})$). However, from visual inspection of the inventories (**Figure 5.9**), Method 2 may not always be most appropriate as the convex hull minimum bounding geometry is always forced towards larger areas (A_{CH}) in order to contain the entire landslide polygon, and thus larger elliptical approximations of that convex hull ($Ef(A_{CH}, P_{CH})$). The larger the area of the elliptical approximation, the greater the probability that there will be a large area of intersection (A_I) between the ellipse and the landslide polygon, and thus a high value of ellipticity. We also find that Method 1 ($Ef(A_L, P_L)$) is particularly sensitive to the level of detail the landslide perimeter has been mapped to, which will vary between inventory producers (E. Harp, personal communication, May 2014), and often results in very long, thin elliptical approximations when there is a high perimeter to area ratio. Consequently, we select Method 3 ($Ef(A_{CH}, P_{CH}) \rightarrow (A_L)$), which is somewhat of a trade-off between lower values of ellipticity, but a more realistic approximation of most landslide polygons.

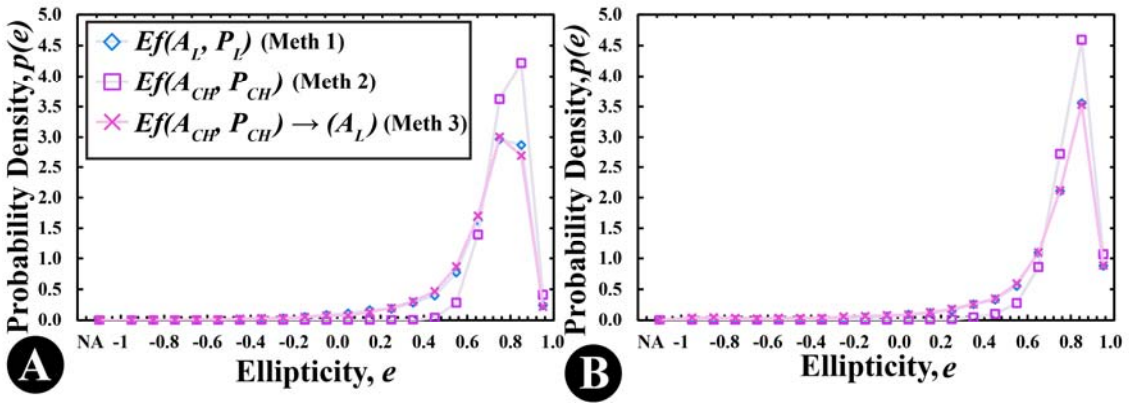


Figure 5.11 *Probability density distributions of goodness-of-fit (measured as ellipticity, e,* ***Eq. 5.4)*** *of elliptical approximations of landslide polygons using Methods 1 to 3 described in* ***Table 5.2*** *for* ***(A)*** *11,111 landslides triggered by the Northridge Earthquake, USA and* ***(B)*** *9,594 landslides triggered by Hurricane Mitch in Guatemala.*

To assess what 'threshold' value of e defines a reasonable approximation of landslide ellipse approximation, a detailed visual inspection of the inventories was performed for both the Guatemala and Northridge landslide inventories. Elliptical approximations were ordered smallest to largest, and a sample of landslides and ellipses were inspected in $e = 0.1$ increments (**Figure 5.10**). From this inspection, we decided the threshold around $e = 0.5$ is a reasonable cut-off between landslides that are well approximated by an ellipse ($e \geq 0.5$)

and those that are not ($e < 0.5$). **Figure 5.12** shows a range of landslide polygons and elliptical approximations for values around this threshold of $e = 0.5$. Although the distinction is not perfect, this does appear to filter out many landslide complexes which may be the result of interacting processes or difficulty in the mapping stage to distinguish between individual landslides. An example of landslides that would be accepted or rejected using the threshold of $e = 0.5$ is shown in **Figure 5.13**.

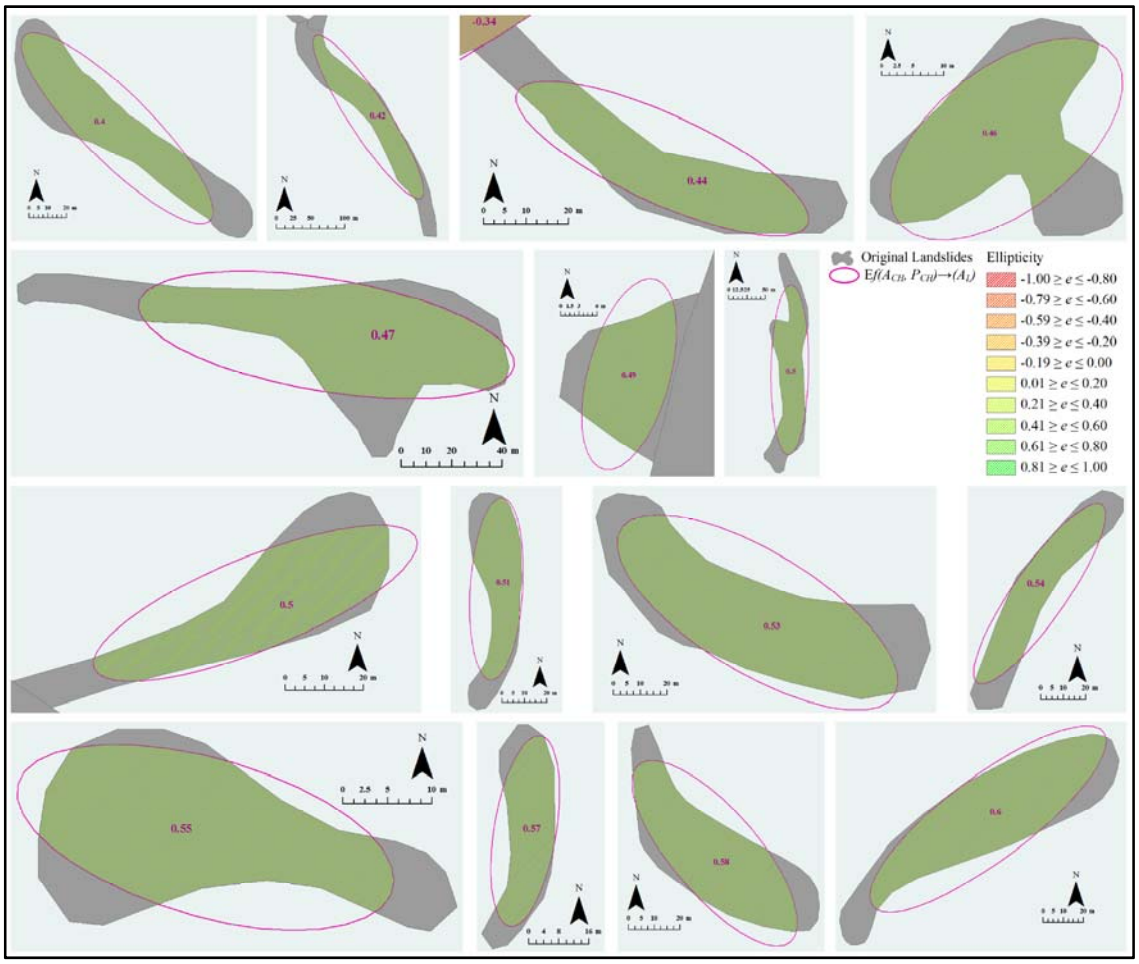


Figure 5.12 *Landslide polygons and elliptical approximations from Method 3 ($Ef(A_{CH}, P_{CH}) \rightarrow (A_L)$,* ***Table 5.3****) and ellipticity (e) measure of goodness-of-fit (****Eq. 5.4****) for a selection of landslides where e = 0.4–0.5. We define landslide ellipses where e ≥ 0.5 to be a reasonable approximation of the original landslide polygon. Landslides from inventory of 9,594 landslides triggered by Hurricane Mitch in Guatemala (Bucknam et al., 2001).*

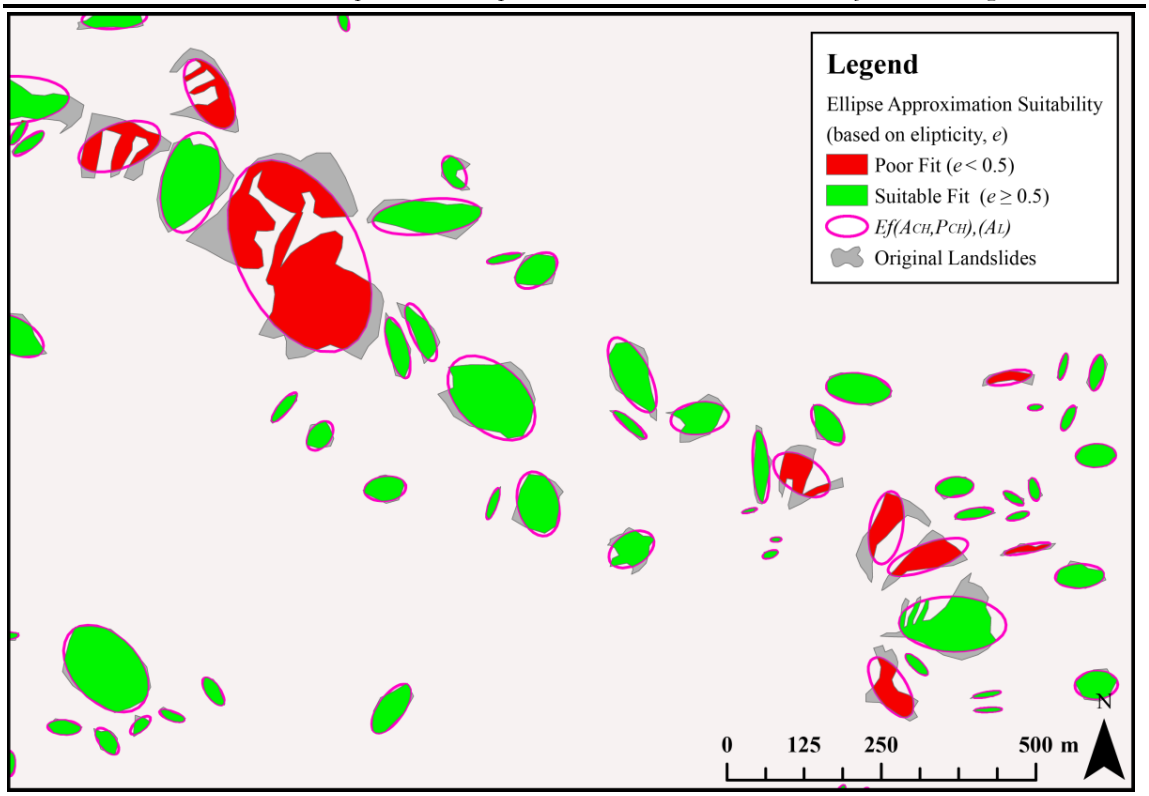


Figure 5.13 *Example of landslides ellipse approximations that would be accepted (green) or rejected (red) based on the criteria of ellipses with ellipticity goodness-of-fit (e) ≥ 0.5 being a reasonable approximation of landslide shapes. Elliptical approximations of real landslides are calculated using Method 3 in* ***Table 5.3*** *(Ef(A_{CH}, P_{CH})→(A_L)). Ellipticity (e) is calculated from the area of the intersection (A_I) between the original landslide polygon area (A_L) and the area of the elliptical approximation (A_E), (**Eq. 5.4**). Landslides from inventory of 11,111 landslides triggered by the Northridge Earthquake (Harp and Jibson, 1995).*

Using Method 3 (E*f*(A_{CH}, P_{CH})→(A_L)) (**Table 5.3**) to calculate elliptical approximations of landslide polygons and rejecting any landslide elliptical approximations where $e < 0.5$, we remove 1,670 landslides (15 %) from the Northridge inventory and 1,736 landslides (18 %) from the Guatemala inventory. We believe that for Northridge, the majority of landslides rejected are large complexes where it was not possible for the geomorphologist to delineate individual landslides. For Guatemala, we believe the majority of landslides rejected are very long (of the order of kilometres), thin debris flows which follow the river channel morphology.

For Northridge, it may have been possible to manually go through complexes and visually separate out individual landslides, although **Figure 5.14A** illustrates that in some cases, almost the entire hillside has failed and it is not clear what represents an individual landslide. For Guatemala, it may be possible to derive another method of calculating the length of long sinuous debris flows, such as the topological skeleton, which reduces shapes to a line equidistant from either boundary of a shape (Supriana and Aryan, 2011). However, after experimentation within a GIS, we found this method to produce inconsistent results

and create problems dealing with branched debris flows, an example of which is shown in **Figure 5.14B**. As we focus on simple and rapidly applicable techniques to approximate landslide shape, we made the decision to reject landslides where $e < 0.5$, acknowledging that this may affect results which we discuss in **Section 5.7**. In **Section 5.5.3**, we show how length-to-width ratios of ellipses are calculated and in **Section 5.5** we analyse statistical trends in ellipse length-to-width ratio.

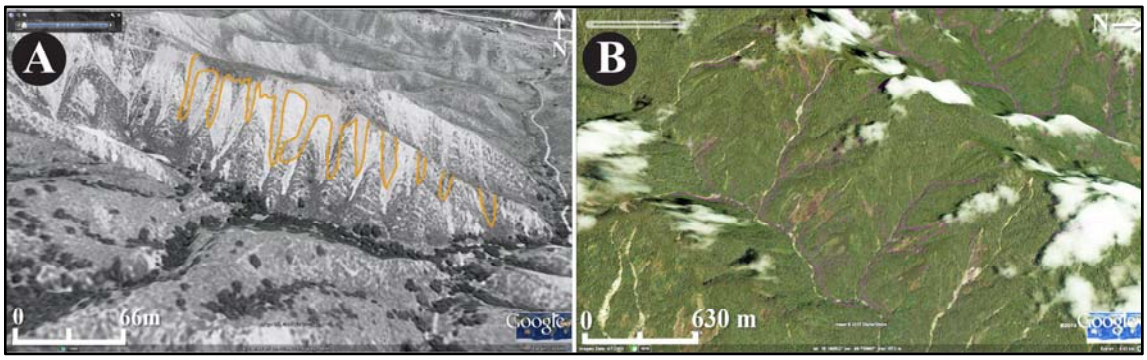


Figure 5.14 *Examples of landslide complexes and debris flows that are removed from analysis draped over Google Earth Imagery and Elevation close to the time of the triggering event.* ***(A)*** *A large complex landslide triggered by the Northridge Earthquake, where is it possible that several smaller landslides have coalesced to create an irregular, branched form that is not well modelled by an ellipse. Google Earth Imagery from 1st June 1994 (~ 5 months after triggering event) (Google Earth, 2014)* ***(B)*** *A large, branched debris flow that is mapped as one polygon, but is most likely the result of coalescence of several run-outs from source areas on different slopes. Google Earth Imagery from April 2003 (~5 years after the event although several scars are still visible) (Google Earth, 2014). Note, in both (A) and (B) there is a slight offset in alignment between the landslide polygons and the imagery, due to conversion to Google Earth Projection and visualisation at an angle. Imagery: Google (2015). Northridge landslide polygons: Harp and Jibson (1995). Guatemala landslide polygons: Bucknam et al. (2001).*

5.5.3 Measuring the Length-to-width Ratio of best fit Ellipses

In **Section 5.5.1** we fit ellipses to landslides and in **Section 5.5.2** we demonstrated how the goodness-of-fit of those ellipses was tested, and 15 % (Northridge) and 18 % (Guatemala) of landslide ellipses removed from analysis due to a poor elliptical approximation. For the remaining ellipses, we calculate the length-to-width ratio (L/W), where L is the longest axis of the ellipse and W is the widest point of the ellipse perpendicular to L. This was shown diagrammatically in **Figure 5.7** (**Section 5.5.1**). **Figure 5.15** shows boxplots of the L/W values for Northridge and Guatemala, showing that the distribution in landslide ellipse length-to-width ratio (L/W) is relatively similar for both the Northridge and Guatemala inventories. When considering both inventories, $\sim 1.2 < L/W < \sim 15$, and the majority of observations (interquartile range) lie between $1.8 < L/W < 3.3$. Median L/W is 2.5 and 2.2 for Northridge and Guatemala, respectively.

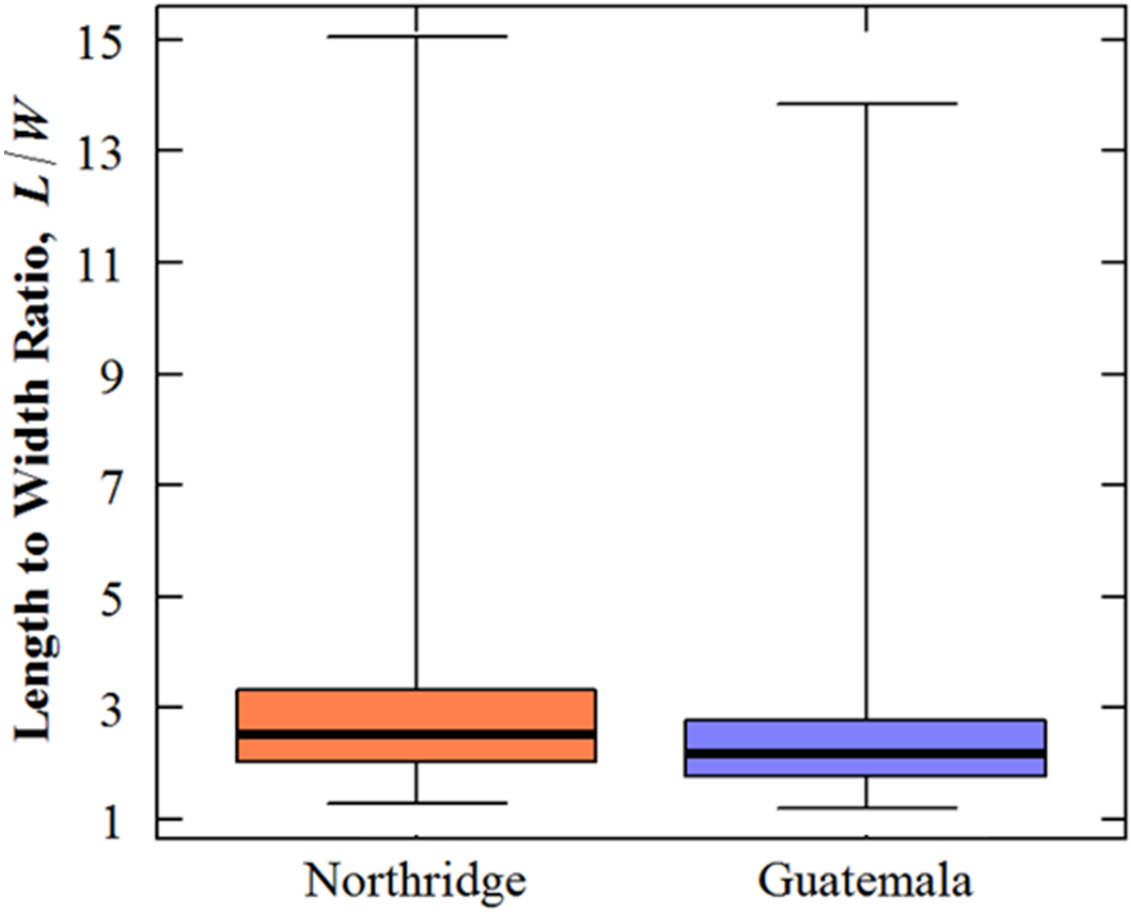


Figure 5.15 *Boxplots of landslide ellipse length-to-width ratio (L/W) for all landslide ellipses where ellipticity (e) ≥ 0.5 for Northridge and Guatemala triggered landslide event inventories.*

Figure 5.16 shows twelve visual examples of L/W, which suggests a potential relationship between L/W and landslide area (A_L), as the larger landslides shown appear to be longer and thinner than smaller landslides. In the following **Section 5.6** we test whether L/W varies with A_L by splitting landslides into area categories (**Section 5.6.1**), examining the statistical distribution of L/W in each category (**Section 5.6.2** and **5.6.3**). In **Section 5.6.4** we apply the methods established in **Section 5.5** and **5.6.1–5.6.3**. to five other 'lower confidence' inventories (described in **Table 5.2**) to examine whether there appear to be generally applicable trends in the statsitical distribution of landslide L/W.

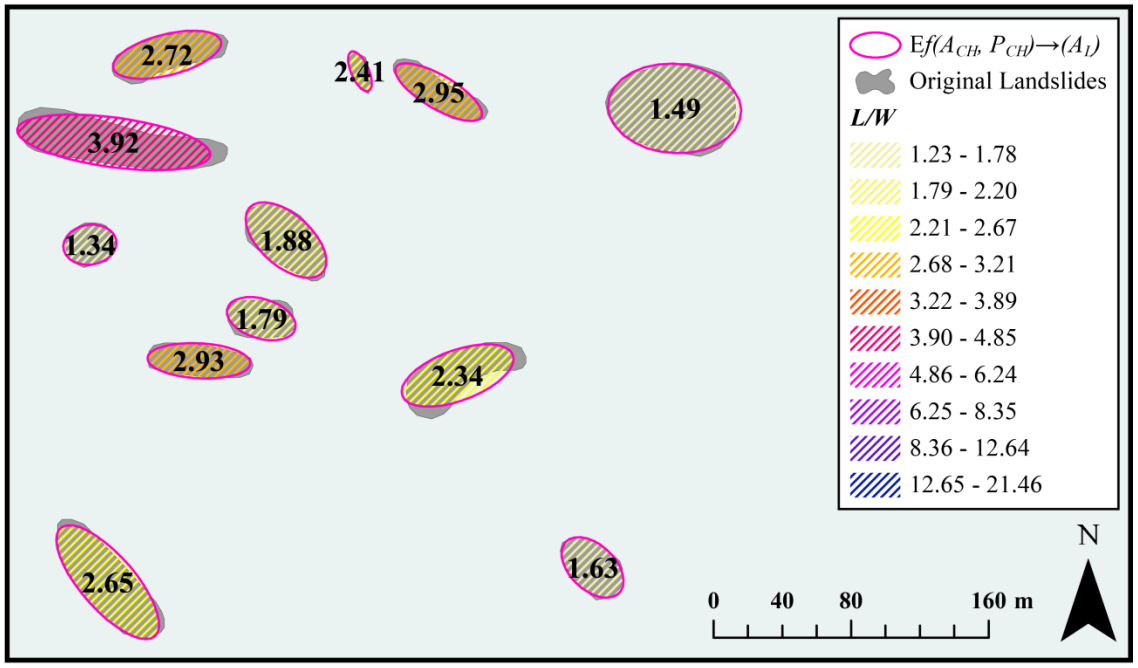


Figure 5.16 *Examples of landslide ellipse length-to-width ratio for twelve ellipses with goodness-of-fit ellipticity value $e \geq 0.5$. Landslides from inventory of 9,594 landslides triggered by Hurricane Mitch in Guatemala (Bucknam et al., 2001).*

5.6 Statistical Analysis Methods and Results

5.6.1 Method to Separate Landslide Ellipse Length-to-width Ratio into Categories of Landslide Area

In **Section 5.5.3**, we hypothesised that there may be a link between the size of a landslide polygon (A_L) and its shape, measured by length-to-width ratio (L/W). To investigate this, we split landslides areas A_L into multiple categories and look at the distribution of L/W in each area category. As landslide areas are known to span several orders of magnitude, and for medium and large landslide areas an inverse-power law distribution (Stark and Hovius, 2001; Malamud *et al.*, 2004a), we chose to split landslides into approximately logarithmically increasing categories of A_L. **Figure 5.17** shows boxplots of L/W by these area categories for both the Northridge and Guatemala inventories (for landslides where $e \geq 0.5$).

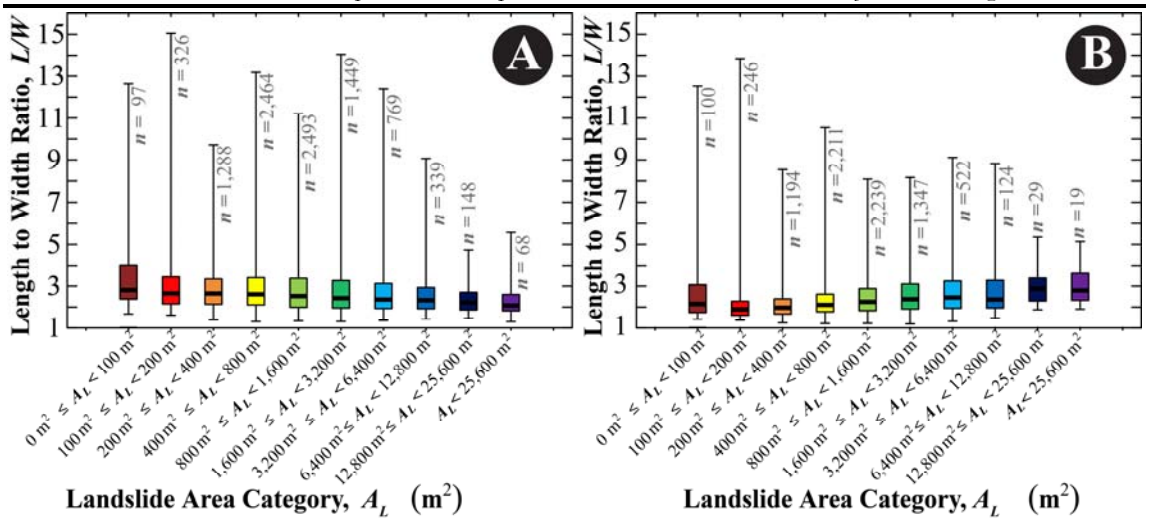


Figure 5.17 *Boxplots of landslide ellipse length-to-width ratio (L/W) separated into categories by landslide area (A_L), which increase approximately logarithmically.* ***(A)*** *9,441 landslides in the Northridge inventory and* ***(B)*** *8,031 landslides in the Guatemala inventory.*

Figure 5.17 shows that for Northridge, landslide ellipse L/W range and median decreases with landslide area (A_L), whereas the opposite is observed in Guatemala. We use a Kolmogorov-Smirnov (K-S) test (discussed in more detail in **Section 5.6.2.3**) to compare the distribution of L/W in each landslide area category, which we will give numbers of 1 to 10 (**Figure 5.17**, left to right along the x-axis), the hypothesis being:

H_0 *The L/W values in landslide area category i come from the same distribution as the L/W values in landslide area category j*, where *i and j represent the ten different categories of landslide areas given in* ***Figure 5.17***, $i \neq j$, $i = 1,2,3,\ldots,10$; $j = 1,2,3,\ldots,10$.

H_1 *The L/W values in landslide area category i do not come from the same distribution as the L/W values in landslide area category j*, where *i and j represent the ten different categories of landslide areas given in* ***Figure 5.17***, $i \neq j$, $i = 1,2,3,\ldots,10$; $j = 1,2,3,\ldots,10$.

The K-S test measures the maximum distance, D between two cumulative frequency curves (in this case this corresponds to the cumulative frequencies of L/W in each A_L category). The maximum distance D is then compared to already established tables to calculate the probability, p that this magnitude of D would have been observed if the two datasets did truly come from the same distribution. **Figure 5.18** shows the K-S test p value results of comparing each landslide area category (CA_L) to each other for Northridge and Guatemala.

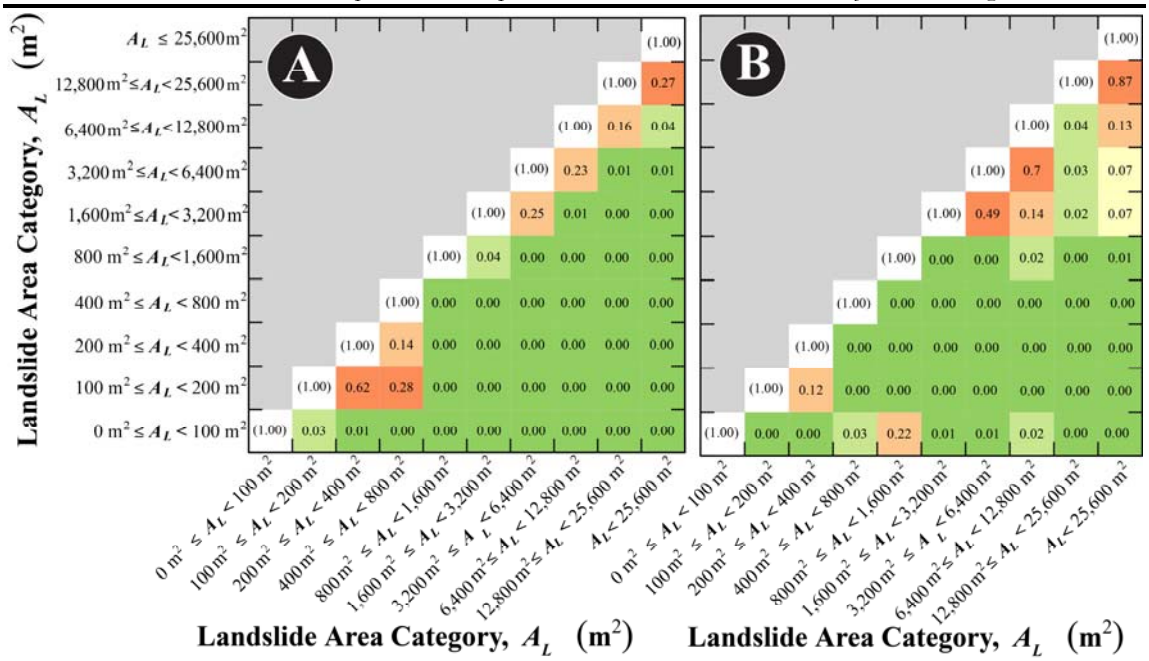


Figure 5.18 *p values from Kolmogorov-Smirnov tests to compare distributions of landslide L/W between each area category.* ***(A)*** *9,441 landslides in the Northridge inventory and* ***(B)*** *8,031 landslides in the Guatemala inventory. Box colours correspond to the significance of the p-value, where green denotes very low p-values (i.e., there is a low probability the samples were drawn from the same distribution) and orange denotes high p-values (i.e., there is a high probability the samples were drawn from the same distribution). Values where p = 0.00 denote very small values of p (e.g., $p = 1 \times 10^{-5}$) that are automatically rounded to 0.00 by R statistical software.*

Using a significance level of p= 0.1, for many A_L categories, we can reject the null hypothesis that the probability density distribution of L/W is the same in each A_L category. The p values are generally higher in neighbouring landslide area categories for the 'tails' of the categories (i.e., very small and very large landslide areas), which we attribute to small sample sizes in these categories. Broadly, we accept that there is a difference in the distribution of L/W between landslide area categories, although not perfect. In the following **Section 5.6.2** we describe methods to fit statistical probability density functions to L/W in each landslide area category.

5.6.2 Methods to Fit Probability Density Functions to Observed Length-to-width Ratio and Test the Goodness-of-Fit

5.6.2.1 Maximum Likelihood Estimation to Fit a Probability Density Function to Observed Length-to-width Ratio

Figures 5.15 and **5.17** suggest that landslide ellipse length-to-width ratio (L/W) may be non-linearly and non-normally distributed, as are many other environmental phenomena (Turcotte and Malamud, 2004). Thus, common methods of fitting probability density functions (pdfs) to observed data such as regression fitting and method of moments are not appropriate, as these assume the minimum and maximum values of the data can be

constrained (Clauset *et al.*, 2009). A more robust and efficient method of curve fitting is Maximum Likelihood Estimation (MLE) (Myung, 2003). MLE works on the assumption that the distributional family is known a-priori, and attempts to find the parameter values of that distribution that are most likely, given a set of observations. This is done by finding the local maxima of the likelihood function. For fitting a one-parameter probability density function ($F(x,\Theta)$) to a set of observations (x_n, $n = 1,2,3,\ldots,N$), this can intuitively be described as the following iterative process:

1. *Starting parameter value(s) are defined* for the distribution $p(\Theta)$.
2. *Probability density values (y) are calculated* for each observation of x.
3. *The probability density function* ($F(x,\Theta)$) *is evaluated at observed values of x,* given parameter value Θ.
4. *The likelihood function* ($f(x|\Theta)$ is the product of all probability densities ($(F(x_1|\Theta) \times F(x_2|\Theta)\ldots \times F(x_n|\Theta))$.
5. *Steps 1-4 are repeated hundreds to thousands of times where* Θ *takes different values*
6. *The optimal value of* Θ will be that that produces the maximum value of the likelihood function ($f(x|\Theta)$).

In practice, it is more common to use the logarithm of the likelihood function, as it is more computationally efficient to deal with summation than multiplication, in which case, the log likelihood function becomes the sum of probability densities $f(x|\Theta)$, rather than the product (White, 2010).

For some simple models such as the above one parameter example $F(x,\Theta)$, maximising the likelihood function (and thus selecting model parameters) may be achieved analytically by solution of a partial differential equation (Myung, 2003). However, for more complex models with a large number of parameters, a local maxima is found iteratively using statistical software (e.g., stats4 or MASS R statistical software packages) (Ricci, 2005). Consequently, both the probability density function and parameters selected may not optimally describe the observed data, depending on the model complexity and number of iterations to solve the likelihood function.

To decide which family of probability density function to fit to observed values of L/W in each landslide area category in the Northridge and Guatemala landslide inventories, we initially used a curve fitting software package (TableCurve2D, 2015) which uses MLE to fit > 3,600 pdfs to the data. Out of > 3,600 distributions, we found that the three-parameter inverse Gamma probability density function (**Eq. 5.5**) ranked highly in terms of goodness-of-fit and minimising he number of parameters used to describe the pdf.

$$p(L_E/W_E) = \frac{1}{a\Gamma\rho}\left[\frac{a}{\frac{L}{W}-s}\right]^{\rho+1} exp\left[-\frac{a}{\frac{L}{W}-s}\right] \quad (5.5)$$

Where:
L_E/W_E = *Length-to-width ratio of ellipse*
ρ = *Shape parameter primarily controlling inverse power law decay (right hand tail)*
a = *Location parameter primarily controlling the position of the maximum probability (rollover)*
s = *Scale Parameter controlling the gradient of the left hand tail of the distribution*
Γ = *Gamma function*

The inverse Gamma pdf was introduced in **Chapter 4** to describe the probability density distribution of landslide area (A_L) (Malamud *et al.*, 2004a, Malamud *et al.,* 2004b). With regard to describing the probability density distribution of L/W, the inverse Gamma pdf models a relatively low probability of observing small values of L/W; it has an exponential rollover for small to medium values of L/W; and it has an inverse power-law decay for large values of L/W. In practical terms, this means there is a relatively low probability of observing very 'compact' ellipse shapes (i.e., L/W approaching 1.0, which is a circle), there is a high probability of observing middle values of L/W and a medium to low probability of observing very long, thin landslides (e.g., $L/W > 10$).

Because the curve fitting software we initially use is somewhat of a 'black box', we then developed codes (**Appendix A**) in R statistical software to robustly fit the a-priori selected inverse Gamma probability density function to each dataset using MLE and test the goodness-of-fit. In the following **Section 5.6.2.2** we describe a bootstrapping technique used in conjunction with the MLE fitting, which allows approximation of the range of uncertainty for the parameter values fit to the data, and present results of fitting the inverse Gamma pdf. In **Section 5.6.2.3** we present methods to test the goodness-of-fit of these pdfs to observations.

5.6.2.2 Bootstrapping Technique to Obtain Uncertainty Estimates of Probability Density Function Parameter Values

Section 5.6.2.1 outlined the maximum likelihood (MLE) method of fitting an inverse Gamma probability density function (pdf) to the observed L/W data, and highlighted the issue that more than one combination of parameter values could describe the distribution of the data well. We somewhat account for this by using a bootstrapping technique to obtain an estimate of the uncertainty range of each parameter value (Efron and Tibshirani,

1993). This involves creating many samples of the data, MLE fitting a pdf to each sample, and investigating the variability in the resulting best fit parameter values.

The principle of the bootstrapping technique is the assumption that the observed data sample is actually the 'population', i.e., the observed data occupies the entire range of potential values (BootAnim, 2015). By assuming this, we then create samples from the data to simulate the variability in the dataset caused by sampling. For our observed L/W dataset of N_{LS} number of observations, the steps are:

1. *Randomly sample N_{LS} values* from the original dataset, with replacement (i.e., the same individual observation can be sampled multiple times).
2. *Fit the pdf to the sample* and record parameter values.
3. *Repeat steps 1-2* a number of times (we choose 1000 iterations).
4. *Investigate variability in parameter values* over all iterations.

Each sample contains the same number of observations (N_{LS}) as the original dataset. Some original values will be sampled multiple times, whereas others may not be sampled at all. Values of L/W around the maximum probability (rollover) of the distribution are more likely to appear in the sample multiple times than values in the left or right tails of the distribution, where there are a smaller number of observations, and thus lower probability of being selected. Resultantly, the sampled dataset should approximately follow the same pdf as the original dataset but with some variability. **Figure 5.19** shows an example of this using the Northridge inventory landslide ellipse L/W values (entire dataset).

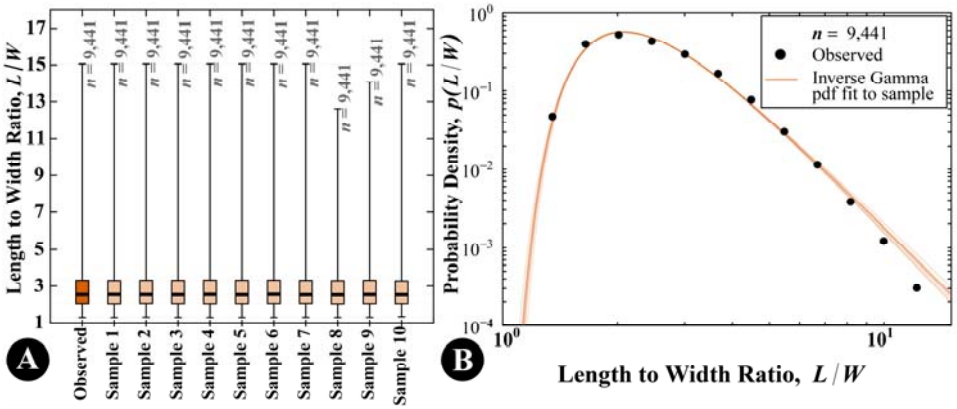
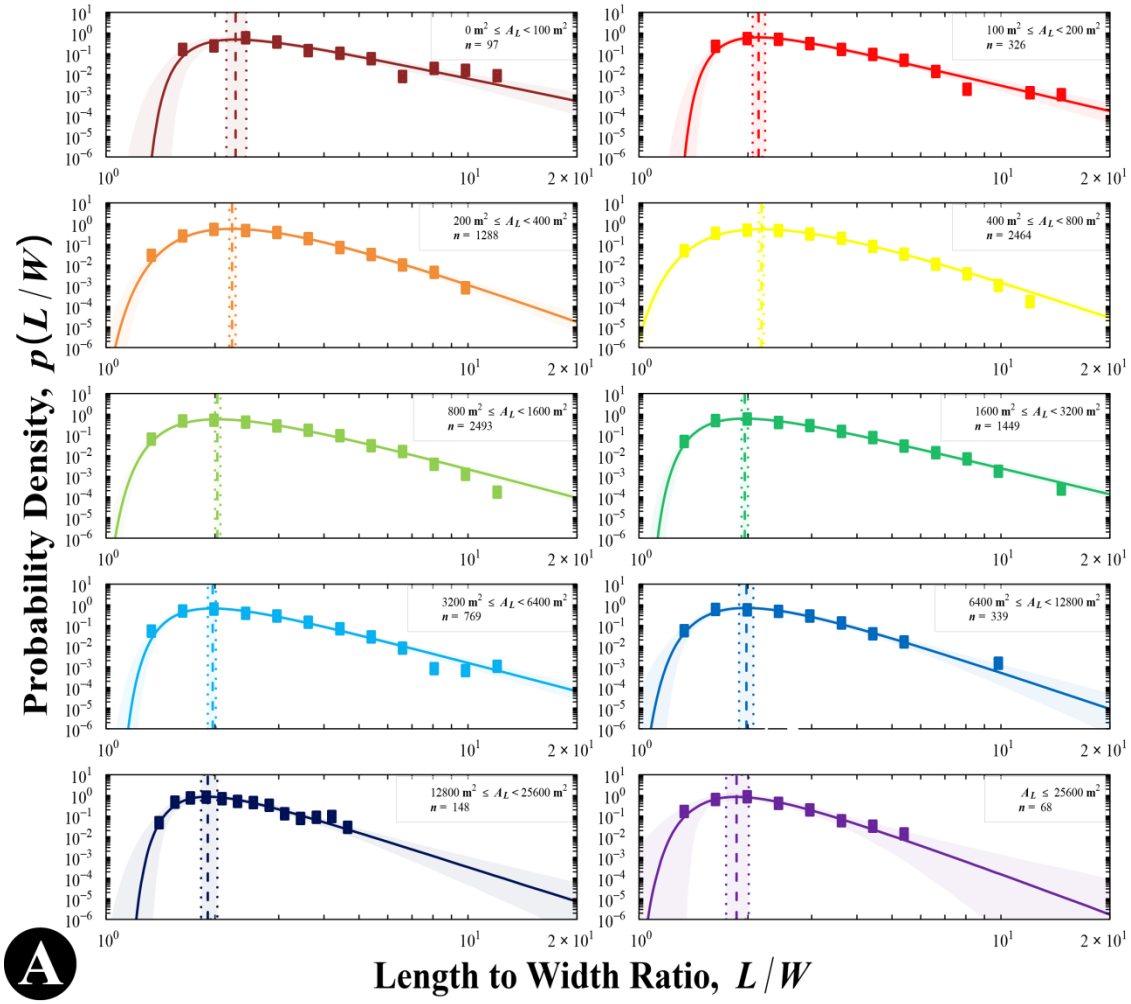


Figure 5.19 *Demonstration of bootstrapping technique to create ten samples of the data to obtain an uncertainty range on pdf parameter values, applied to all length-to-width (L/W) values for areas from Northridge landslide inventory, where L/W is determined using the Ef($(A_{CH}, P_{CH})\rightarrow A_L$) technique (**Section 5.5.1**). **(A)** Boxplot showing original data and ten samples drawn from the original data. Sample size is the same size as the original dataset, samples are drawn with replacement. **(B)** Inverse Gamma pdfs fit to each sample shown in (A). Due to each sample being slightly different, the best MLE pdf fit differs each time. Pdfs lines are semi-transparent, so the darker the line colour, the more pdfs are overlapping.*

In **Figure 5.19A**, a boxplot of the original observed values of L/W is shown next to ten samples of the data (sample size $n = N_{LS}$), and we observe that the distribution is approximately the same, although does not extend to the full range of the observed data in each case. In **Figure 5.19B**, we show the pdfs fit to each of those ten samples, overlaid with the probability densities of the observed data. We observe that there is greater agreement in the pdfs around the rollover of the distribution and less agreement in the tails of the distribution. This is due to fewer observations in the tails of the distribution, meaning a lower probability of being sampled in each iteration of the bootstrap.

In Figure 5.20 we show the results of MLE fitting (**Section 5.6.2.1**) the inverse Gamma pdf to L/W data (where $e \geq 0.5$) for each landslide area category and bootstrapping to obtain an uncertainty range around the pdf fit for the Northridge and Guatemala inventories. Observed values of probability density are also shown. Probability density is calculated using approximately logarithmically increasing L/W bin sizes.



*(**Figure 20** continues on next page)*

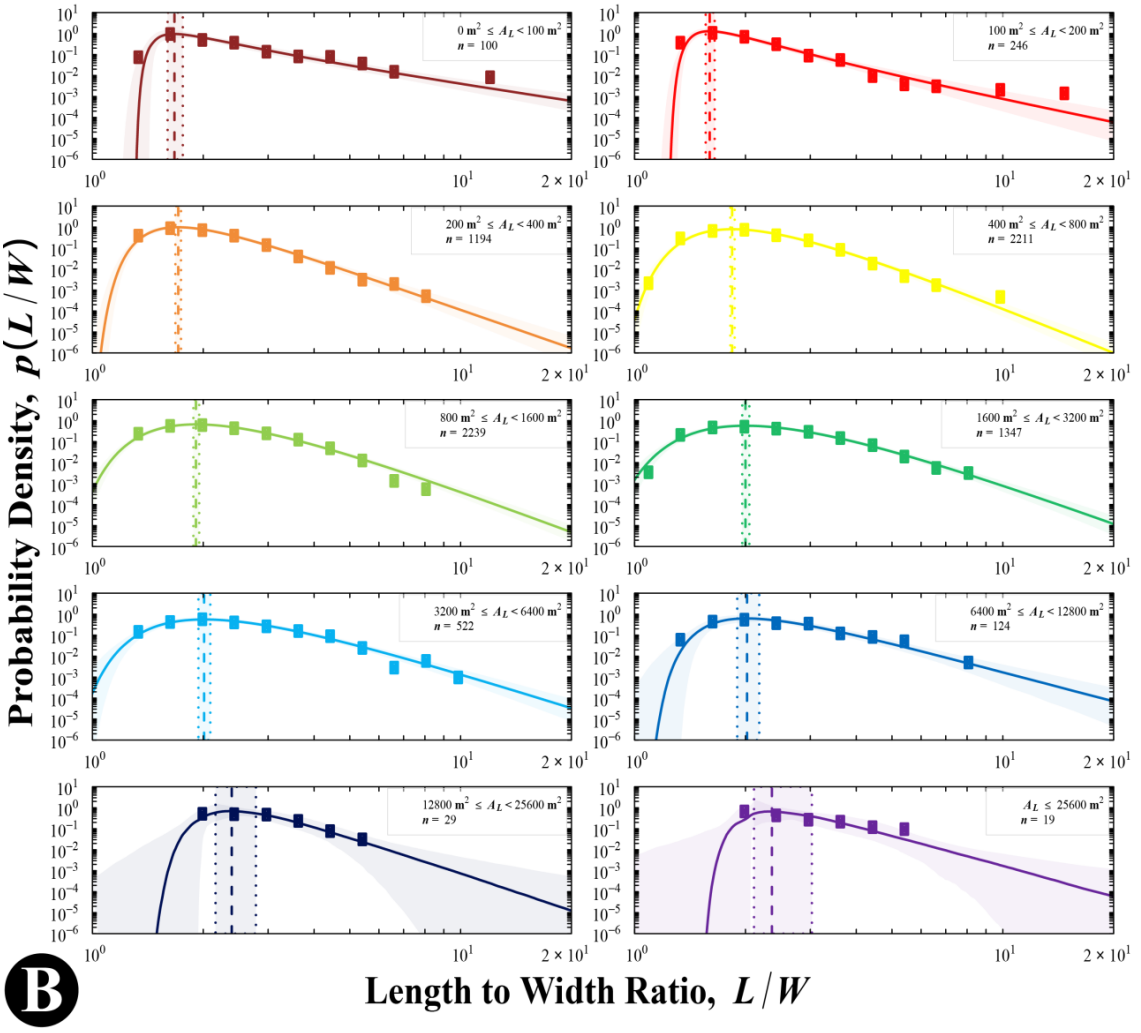


Figure 5.20 *Inverse Gamma probability density function (**Eq. 5.5**) fit to landslide ellipse length-to-width ratios (L/W) using maximum likelihood estimation (**Section 5.6.2.1**) where landslides are split into categories based on the landslide area (A_L). (**A**) 9,441 landslides triggered by the 1994 Northridge Earthquake, California USA (**B**) 7,858 landslides triggered by 1998 Hurricane Mitch in Guatemala. Observed values of probability density are also shown. The shaded area either side of the pdf line represents a form of 5th/95th percentile confidence intervals around the pdf. These are calculated using a bootstrapping technique (**Section 5.6.2.2**) where observed data is repeatedly sampled with replacement and an inverse Gamma pdf fit to each sample. The location of maximum probability (rollover) is shown as a dashed line and the 5th/95th percentile values of rollover from the bootstrapping technique are shown as the shaded vertical area. Points represent observed probability densities, which are calculated using approximately logarithmically increasing bin sizes.*

Figure 5.20 demonstrates that the inverse Gamma pdf appears to describe the probability density distribution of landslide ellipse *L/W* in each landslide area category well for both Northridge and Guatemala. For both inventories, the pdfs fit to the smallest and largest landslide area categories show a wider range of uncertainty (i.e., the 5th and 95th percentile pdfs are quite far apart), which we attribute to small sample sizes in these categories. In the following **Section 5.6.2.3** we present a method to test the goodness-of-fit of the pdfs to the observed data.

5.6.2.3 Monte Carlo technique to test goodness-of-fit of pdf fit to observed data

To test whether the inverse Gamma pdf fit to observed landslide ellipse L/W values for Northridge and Guatemala inventories is statistically significant, we look at how far observed values deviate from the distribution fit to them. We then assess the probability that if the observations were truly from that distribution, that the deviation between observation and distribution would be the size it is.

As the observations of L/W are not normally distributed, not all goodness-of-fit tests are suitable. Here, we use a modified version of the Kolmogorov-Smirnov (K-S) test to measure the maximum difference between the observed cumulative frequency distribution and the theoretical cumulative frequency (i.e., the fit distribution) (James, 2006). The K-S test was selected as it is relatively intuitive and due to its non-parametric nature is simple to apply to the inverse Gamma pdf. The test is also suitable for small sample sizes and is considered to be a 'powerful' statistical test (Lillefors, 1967). This test is most sensitive to the peak of the distribution (where most data points lie).

In its standard form, the K-S test measures the maximum distance (D) between two cumulative frequency curves: (i) the cumulative frequency of the observed (empirical) data ($F_n(x)$) and (ii) the cumulative frequency of the distribution fit to the data ($F(x)$). If the maximum distance (D) between two curves is low, this suggests both datasets ($F_n(x)$ and $F(x)$) follow the same distribution. If D is high, this suggests significant deviation between the cumulative distribution functions of the two datasets. The Kolmogorov-Smirnov test statistic D given by (Wilcox, 2005):

$$D = \sup_x |F(x) - F_n(x)| \qquad \textbf{(5.6)}$$

Where:
sup = supremum (maximum) value of the set of distances between observed and fit cumulative distributions (|F(x) – Fn(x)|) evaluated for each observed value of x
$F_n(x)$ = Empirical cumulative distribution for each observed value of x
F(x) = Theoretical 'fit' distribution evaluated at each observed value of x

Figure 5.21A shows the inverse Gamma cumulative distribution function fit to Northridge data and observed L/W cumulative densities for Northridge landslides in area category 200 $m^2 \leq A_L < 400$ m_2. **Figure 5.21B** shows $F(L/W)$ - $F_n(L/W)$ for observed values of L/W for Northridge (200 $m^2 \leq A_L < 400$ m^2), showing the maximum deviation (D) between the cumulative distribution function and observed cumulative densities, and the L/W at which this occurs.

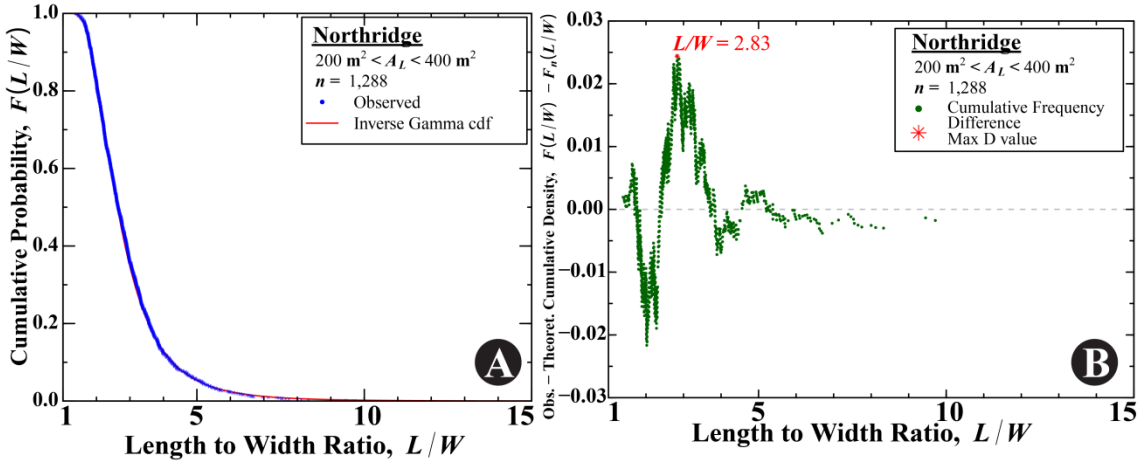


Figure 5.21 *Example of cumulative distribution function (CDF) fit to observed data and location of the maximum deviation between the CDF and observed cumulative density for landslides in the area category 200 m² ≤ A_L < 400 m² from the Northridge inventory.* ***(A)*** *Cumulative distribution function* ***(B)*** *Observed cumulative density (F(L/W)) minus cumulative density function (F_n(L/W)) and location of the maximum deviation (D-value).*

The *D* value is then compared to critical values at significance levels from published tables (calculated using Monte Carlo simulations much like those discussed in **Section 5.6.2.2**). If the *D* value is smaller than the critical value, the null hypothesis can be accepted:

H_0 *The observed data come from the distribution.*

H_1 *The observed data do not come from the distribution.*

The standard K-S test is not appropriate if the parameters describing the distribution have been estimated from the dataset, as the significance levels stated in standard tables will be more conservative than reality (Lillefors, 1967; Crutcher, 1975) (i.e., the probability that the data was observed by chance will be stated as lower than it actually is). Instead, the test can be adapted in a Monte-Carlo type workflow to simulate what typical deviations are observed between a known distribution and random values selected from that distribution. We adopt the process outlined by Clauset *et al.* (2009):

1. *Fit the inverse Gamma probability density function to the observed L/W data* using MLE. This gives us parameter values ρ_{obs}, a_{obs}, s_{obs} describing the best-fit pdf to the data.
2. *Perform a 'standard' K-S test* to calculate the *D* value between the observed cumulative distribution function and the inverse Gamma cumulative distribution function fit to the data. This is D_{obs}.
3. *Randomly generate n_{obs} values from the inverse Gamma pdf* with the parameters ρ_{obs}, a_{obs}, s_{obs}. The variable n_{obs} = the number of values in the observed dataset.
4. *Using MLE, fit the inverse Gamma distribution to these randomly generated values*, this gives a pdf with parameters ρ_{sim}, a_{sim}, s_{sim}.

5. *Calculate the K-S D statistic* between the randomly generated values and the distribution fit to these values. This is D_{sim}.
6. *Repeat* ***Steps 3-5*** for n_{iter} iterations where n_{iter} is related to the desired level of precision (given in the following **Eq. 5.7**). In each case, the values ρ_{sim}, a_{sim}, s_{sim} will vary slightly. For each iteration, record D_{sim}.
7. *Compare D_{obs} to the range of D_{sim}.* The significance level (p) is defined by the percentile of D_{sim}, i.e., to test the data to $p = 0.9$ significance level, compare D_{obs} to the 90th percentile of D_{sim}. If $D_{obs} < p \times D_{sim}$, we can accept the null hypothesis that the observed L/W data come from the inverse Gamma pdf with parameters ρ_{obs}, a_{obs}, s_{obs}.

This method works on the premise that the observed data can be considered to have come from the same distribution as that fit to it if the D value lies within the normal range of those encountered when fitting a distribution to random deviates from a known distribution.

The number of iterations of steps 3-5 (n_{iter}) is defined by the desired precision of the output. Based upon empirical observation, Clauset *et al.* (2009) suggest the number of iterations n_{iter} (for steps 3-5 above) is given by:

$$n_{iter} = \frac{1}{4}\epsilon^{-2} \qquad (5.7)$$

Where:
n_{iter} = number of iterations of the Monte-Carlo K-S goodness-of-fit test
ϵ = the number of decimal places of precision for the p-value.

Figure 5.22 shows boxplots of the distribution of D_{sim} values and D_{obs} for each landslide area category for both Northridge and Guatemala.

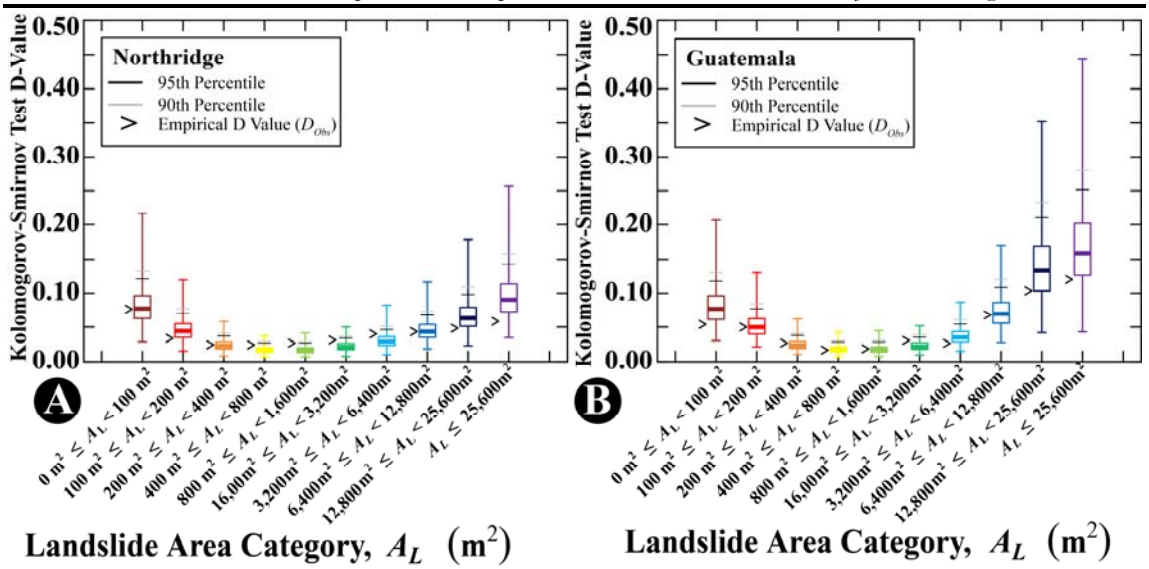


Figure 5.22 *Boxplots of D values from Monte-Carlo Kolmogorov-Smirnov goodness-of-fit testing, compared to observed D value (D_{obs}).*

For all landslide area categories for both Northridge and Guatemala in **Figure 5.22**, D_{obs} is below the 90th and 95th percentile of D_{sim} values (corresponding to p = 0.10 and p = 0.05), suggesting the deviations between observations and the fit pdf are within the 'normal range' to be expected if the data were drawn from that distribution. Thus, we can say that the inverse Gamma pdfs shown in **Figure 5.22** reasonably well model the distribution of landslide ellipse L/W for both Northridge and Guatemala.

5.6.3 Variation in inverse Gamma Probability Density Function Parameters with Landslide Area

Section 5.6.1 showed that the inverse Gamma probability density function (pdf) appears to be a statistically reasonable fit to describe the distribution of landslide ellipse length-to-width ratio (L/W), when landslides are separated into landslide area categories that increase in size approximately logarithmically. In this **Section 5.6.2** we discuss how the inverse Gamma pdf appears to vary with landslide category in terms of visual appearance (**Figure 5.23**) and parameter values (**Figure 5.24**).

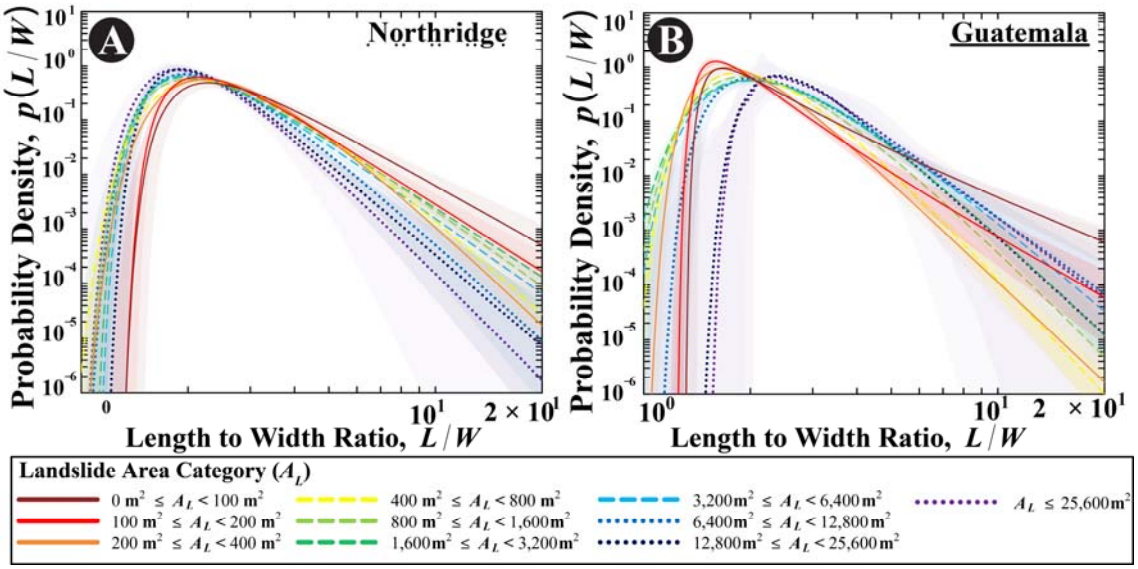


Figure 5.23 *Inverse Gamma Probability Density Functions (pdfs) fit to landslide ellipse length-to-width ratio (L/W) values in each landslide area category.* ***(A)*** *9,441 landslides in the Northridge inventory and* ***(B)*** *8,031 landslides in the Guatemala inventory. The shaded area either side of the pdf line represents a form of 5th/95th percentile confidence intervals around the pdf. These are calculated using a bootstrapping technique (**Section 5.6.2.2**) where observed data is repeatedly sampled with replacement and an inverse Gamma pdf fit to each sample.*

Figure 5.23 shows the inverse Gamma pdf fit to L/W distributions in each landslide area category for Northridge and Guatemala. In the Northridge inventory (**Figure 5.23A)**, we observe that as A_L increases, the distribution shifts towards smaller values of L/W and the gradient of both tails broadly increases. Broadly speaking, this implies that in Northridge, larger landslide areas tend towards more compact shapes with smaller values of L/W, although there is still a spread of L/W values in each area category. The opposite is observed in the Guatemala inventory (**Figure 5.23B**); as A_L increases, the inverse Gamma pdf shifts to larger values of L/W. For medium landslide area categories, the pdf appears to flatten, and thus the gradient of the right hand tail decreases. For the three largest landslide area categories, the gradient of right hand tail of the distribution again increases. Generally, for Guatemala, we observe that as landslide area increases, there is a greater probability of observing longer and thinner shapes (i.e., larger values of L/W). In **Section 5.7**, we discuss that this effect may be more pronounced in Guatemala than shown in **Figure 5.23B**, as many landslide polygons with low values of ellipticity (e) that were removed from analysis are large, long, thin debris flows that have become channelized (**Section 5.5.2**).

Figure 5.24 shows boxplots of the parameter values (ρ, a, s) describing the shape of the inverse Gamma pdf fit to L/W in each landslide area category. The range of parameter values for each location and area category are derived from the bootstrapping technique described in **Section 5.6.2.2**. We observe roughly similar behaviour in parameter values for

both the Northridge and Guatemala inventories, and opposite behaviour between the inventories in terms of the location of the rollover of the distribution.

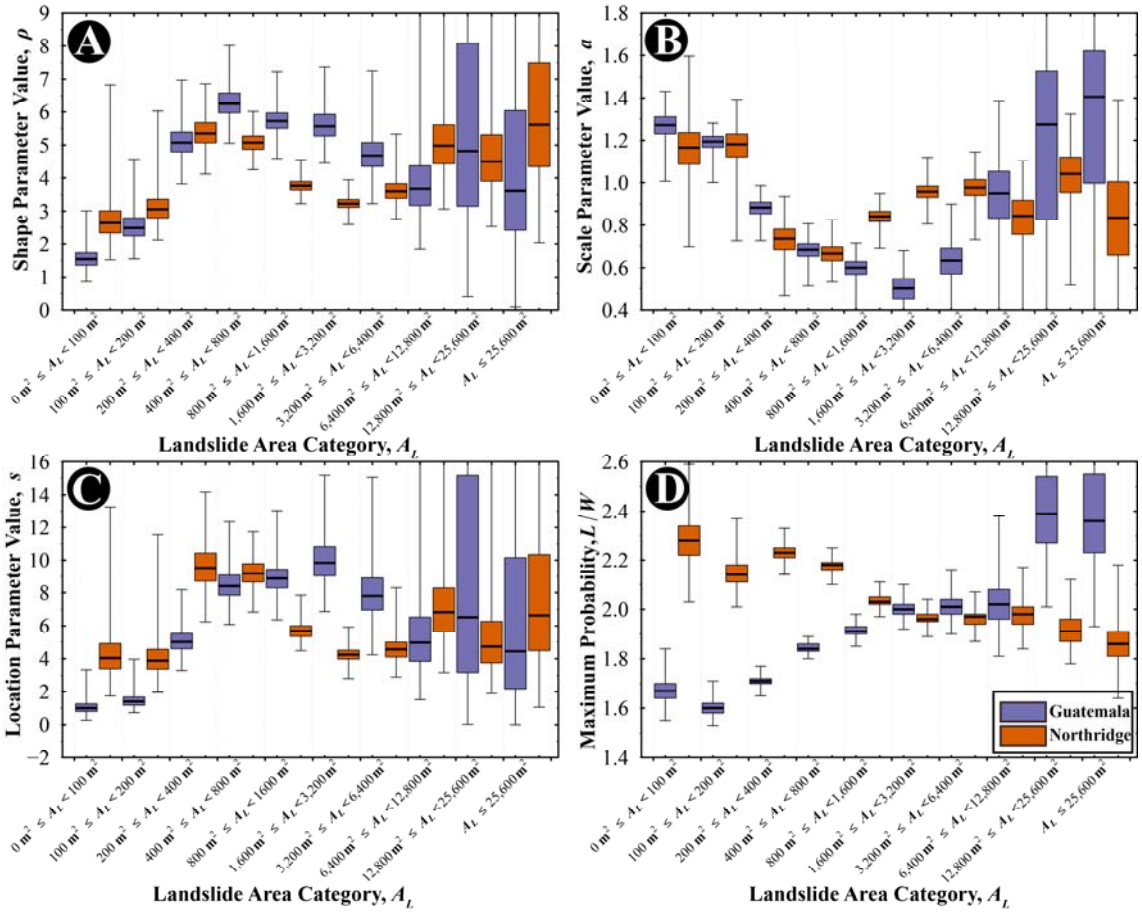


Figure 5.24 *Parameter values and characteristics describing the inverse Gamma pdf fit to landslide ellipse L/W values in each landslide area category for 9,441 landslides in the Northridge inventory (orange) and 8,031 landslides in the Guatemala inventory (purple).* ***(A)*** *Shape parameter (ρ) describing the gradient of the power law decay for medium to large values of L/W.* ***(B)*** *Scale parameter (a) primarily controlling the position of the maximum probability (rollover).* ***(C)*** *Location Parameter controlling the gradient of the left hand tail of the distribution.* ***(D)*** *Location of the 'rollover' of the inverse Gamma pdf, which denotes the value of L/W at which the maximum probability density occurs.*

In **Figure 5.24A**, the shape parameter (ρ) which controls the gradient of the power law decay right hand tail of the distribution is investigated. The variable ρ first increases with A_L, peaking at landslide areas 200 m^2 – 800 m^2. After this point, ρ decreases and then increases again for the two largest landslide area categories. In terms of behaviour in L/W, this trend equates to the right hand tail of the pdf first flattening as A_L increases, meaning that there is a greater likelihood of observing longer, thinner landslides with increasing landslide area (A_L). For landslide areas 800 m^2 ≤ A_L < 12,800 m^2, the gradient of the right hand power law decay then decreases with increasing A_L, so as landslide area increases, there is a greater probability of observing longer, thinner landslides. For the two largest

landslide area categories ρ is relatively high, but also occupies a broader range of potential values from the bootstrapped sampling process (**Section 5.6.2.2**). This greater variability in parameter values can be primarily attributed to a smaller sample size, and greater variability in the distributions when performing bootstrapping MLE technique. **Figure 5.17** showed that the actual distribution of L/W is actually fairly narrow for the largest landslide area categories. But, because there are relatively few observations in these largest A_L, the effect of one 'end member' of L/W being sampled multiple times will have a greater effect in skewing the distribution in a particular direction, and the resulting parameter values from MLE. Consequently, some caution must be exercised in over-interpreting the results of parameter fitting to these largest landslide area categories without larger sample sizes.

Although the overall behaviour in ρ values appear to be the same for Northridge and Guatemala, the actual values of ρ differ between the inventories; ρ being generally higher in Guatemala than Northridge for landslide areas up to approximately 6,400 m^2. After the point, ρ is generally lower for Guatemala than Northridge.

Figure 5.24C illustrates variability in location parameter (s), which primarily controls the exponent of the left hand tail of the inverse Gamma pdf. Overall, the behaviour of s is quite similar to behaviour of ρ, first increasing with A_L up to landslide areas of around 800 m^2, and then decreasing with increasing A_L for landslide areas 800 $m^2 \leq A_L <$ 12,800 m^2, after which point s then increases again. In terms of L/W behaviour, this suggests initially as landslide area increases, the gradient of the left hand tail increases, and thus there is a lower probability of observing low values of L/W. For landslides larger than approximately 800 m^2, s then decreases, so as A_L increases, the decay of the left hand tail also decreases, and thus there is a greater probability of observing L/W values closer to 1 (i.e., very compact landslide shapes).

Figure 5.25 demonstrates the various combinations of ρ and s parameters and their effects on both the left and right hand tails of the inverse Gamma pdf (as the illustrations are hypothetical, the rollover is at the same location). Broadly, we see that for the smallest landslide areas ($A_L <$ 200 m^2), there is a low probability of observing a compact landslide, and a relatively greater probability of observing a long thin landslide. As we move to medium sized landslides (~200 m^2 – 1,600 m^2), there is a relatively higher probability of observing a compact landslide and a lower probability of observing a long thin landslide. Then, as we move to larger landslide areas, landslide L/W occupies a broader range of

values. For the very largest landslide areas, there is a relatively high probability of observing long, thin landslides.

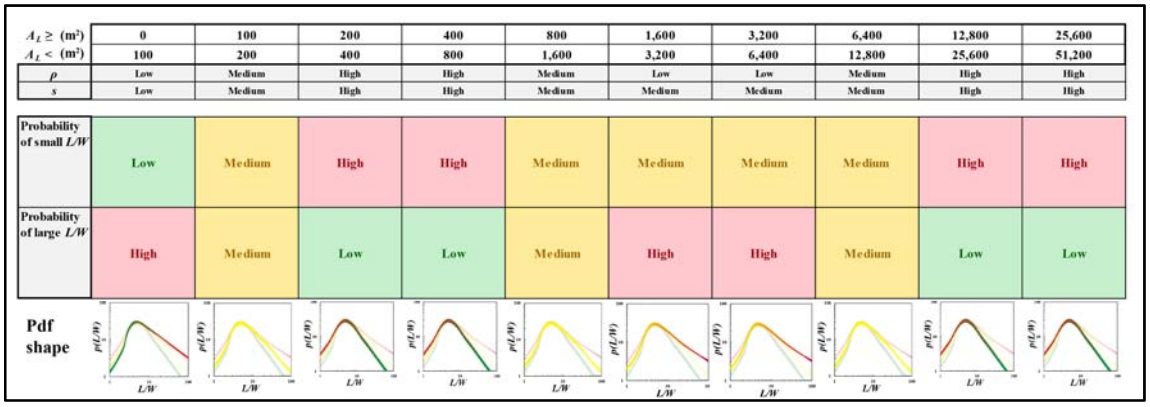


Figure 5.25 *Effects of shape (ρ) and location (s) parameters on the tails of the inverse Gamma probability density function for each landslide area category. Example inverse Gamma pdfs are shown to illustrate differences in the gradients of the tails of the distribution. Parameter values are approximated simply as 'low', 'medium' or 'high' to aid the visualisation of the interaction between parameter values and its effect on the shape of the pdf.*

Figure 5.24D shows the value of L/W at which the maximum probability (rollover) occurs in the inverse Gamma pdf for each landslide area category in both Northridge and Guatemala. The location of the rollover is primarily governed by the scale parameter (a), shown in **Figure 5.24B**, although the location parameter (s) does play a part in the location of the rollover. Here, we observe inverse behaviour between the inventories; for Northridge, the rollover decreases with increasing A_L, whereas for Guatemala, the rollover increases with increasing A_L. This is in agreement with **Figure 5.23**, and we hypothesise that this difference in behaviour can be attributed to differences in the types of landslides in each inventory. The Guatemala inventory is landslides triggered by heavy rainfall in Hurricane Mitch, many of which mobilised into debris flows (Bucknam *et al.*, 2001). Whereas, the Northridge inventory is landslides triggered by an earthquake. This inventory was more dominated by shallow falls and slides, which generally did not mobilise into long-run out landslides (Harp and Jibson, 1995). Indeed, Parise and Jibson (2000) note that mean length-to-width ratios for individual landslides is $L/W = 2.6$, whereas mean length-to-width ratios for landslide complexes (which tend to be larger in area due to coalescence) is $L/W = 1.2$, as many complexes extended along the ridgelines, making them relatively wide in proportion to their length. In the following **Section 5.6.4** we perform the analysis on several other inventories that we have access to, so that we can test whether similar patterns in rollover location are observed. We consider these inventories to be 'lower confidence', either because of less documentation and certainty about how they were

produced (the implications of which is discussed in **Chapter 2**) or because there are relatively few observations, making statistical analysis less robust.

5.6.4 Application to 'Lower Confidence' Inventories

5.6.4.1 Methods

In **Section 5.6.3** we demonstrated that there appear to be some similarities in the trend in inverse Gamma probability density function parameter values (ρ, a, s) describing the distribution of landslide ellipse L/W values for both 9,441 landslides in the Northridge inventory and 8,031 landslides in the Guatemala inventory. However, the overall behaviour of L/W in the two inventories differs; in Northridge, landslides tend towards lower values of L/W with increasing landslide area category, whereas in Guatemala, landslides tend towards higher values of L/W with increasing A_L. We hypothesise that this may be due to the type of landslides which were triggered in each inventory, which is itself a product of geology, trigger mechanism and terrain, land cover and other factors. To further explore these trends in L/W, we now apply the methods outlined in **Sections 5.5** and **5.6.1 – 5.6.3** to five additional inventories (Collazzone, El Salvador, Liguria, Roccafluvione and Taiwan, all described in **Table 5.2**). For each inventory we:

1. *Fit ellipses to landslide polygons* using Method 3 (E*f*(A_{CH}, P_{CH})→(A_L))(**Table 5.3**, **Section 5.5.1**).
2. *Extract only those polygons where an ellipticity goodness-of-fit ellipticity* $e \leq 0.5$ (**Eq. 5.4**, **Section 5.5.2**).
3. *Measure the length-to-width ratio (L/W)* of suitable landslide ellipses (**Section 5.5.3**).
4. *Separate landslides into categories depending on their area* (A_L) (**Section 5.6.1**).
5. *Use Maximum Likelihood Estimation (MLE) to fit an inverse Gamma probability density function (pdf)* to L/W in each CA_L (**Section 5.6.2.1**).
6. *Test the goodness-of-fit* of each inverse Gamma pdf to the observed data using Monte-Carlo Kolmogorov-Smirnov goodness-of-fit testing (**Section 5.6.2.3**).
7. *Explore variability in the inverse Gamma pdf* fit to each landslide area category using a bootstrapping technique (**Section 5.6.2.2 and 5.6.3**).

As discussed in **Chapter 2** and **Section 5.4**, there are very few triggered landslide event inventories that are well documented, substantially complete and contain enough observations to perform robust statistical analysis when landslides are separated into area categories. We acknowledge from the outset of this **Section 5.6.4** that the additional

inventories used for analysis in this section are 'lower confidence' for one or a combination of those reasons, and thus results must be treated with caution. The inventories are described in **Section 5.4**, **Table 5.2**, with notes about why these are lower confidence. Having performed the above bullet pointed steps for each inventory to calculate L/W and fit the inverse Gamma pdf, we now examine similarities and differences between each inventory. We include results from the Northridge and Guatemala inventories for comparison.

Figure 5.26 shows a boxplot of the overall distribution of L/W for all values of A_L for each inventory. Minimum values of L/W ranges 1.2 – 1.5, median L/W ranges 1.9 – 3.2 and maximum L/W ranges 4.2 – 15.1. We next separate L/W in each inventory by landslide area category.

5.6.4.2 Distribution of Length-to-width Ratio for Additional Inventories and Classification into Types of Inventory

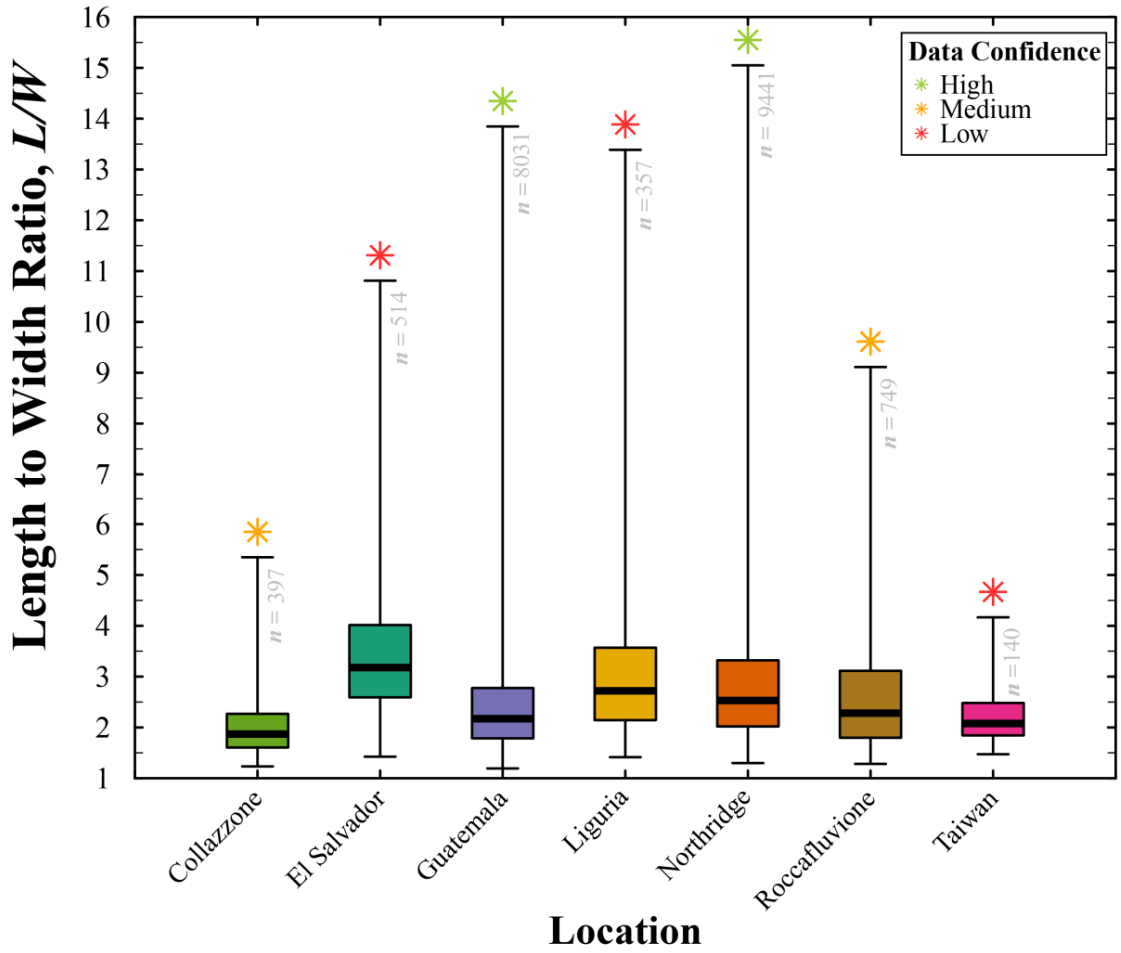


Figure 5.26 *Boxplot of landslide ellipse L/W for all landslide areas in each triggered event landslide inventory (described in Table 5.2). The 'confidence' relating to data quality and suitability is ranked low, medium and high (denoted by the coloured * above each boxplot).*

Figure 5.27 shows boxplots of L/W by landslide area category for each inventory. In this figure, we have arranged inventories by similarity in L/W behaviour:

Column 1 denotes inventories where the increase in L/W with A_L is pronounced (Guatemala, El Salvador and Roccafluvione)

Column 2 denotes inventories where L/W increases with A_L apart from the largest A_L categories and the range of L/W is generally low (Collazzone, Taiwan and to an extent, Liguria)

Column 3 denotes inventories where L/W decreases with A_L (Northridge).

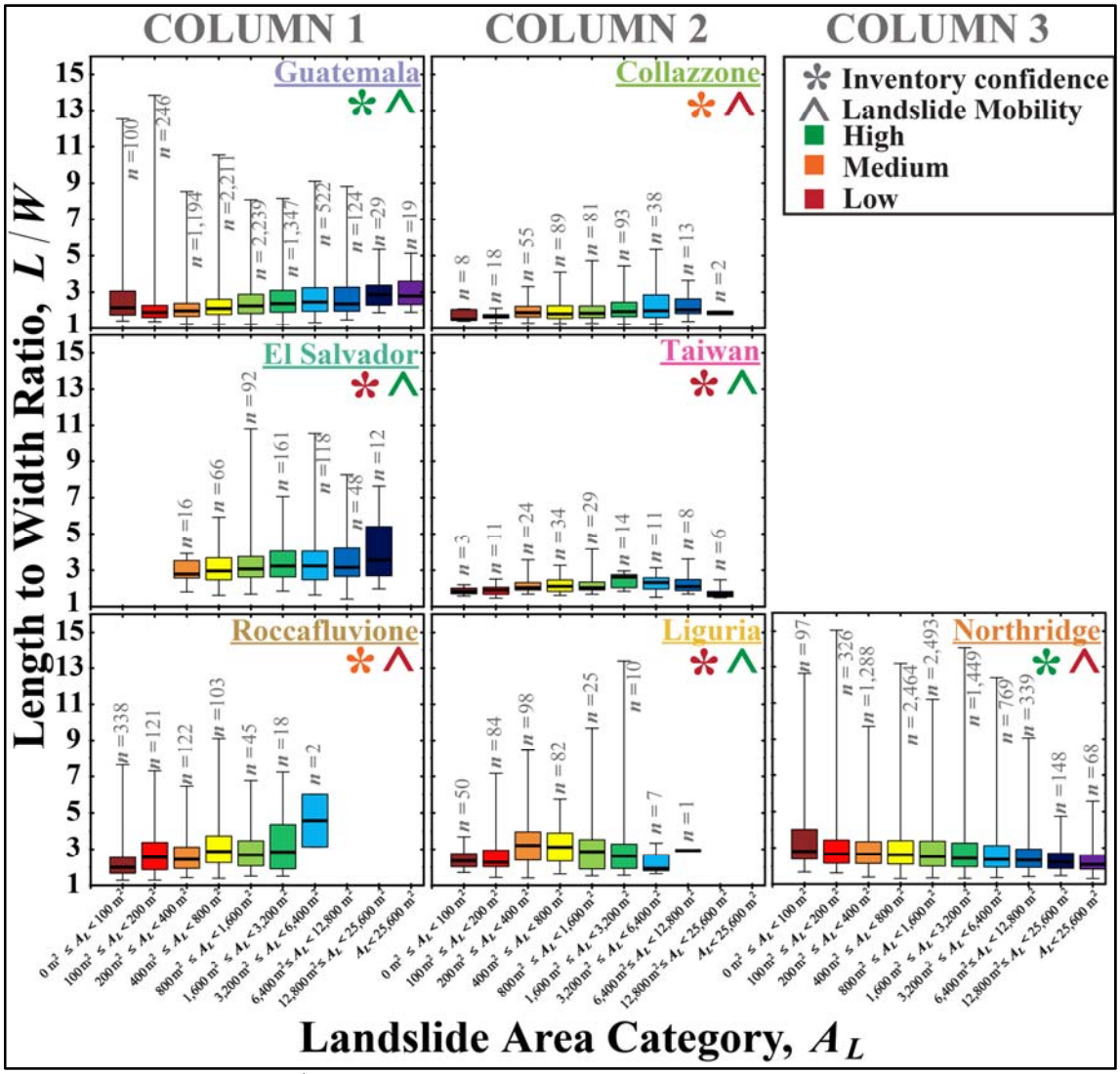


Figure 5.27 *Boxplots of L/W separated by landslide area category for all inventories described in Table 5.2. Landslide area categories increase approximately logarithmically. We sort inventories into three columns, relating to the behaviour of L/W with A_L.* ***Column 1*** *denotes inventories where L/W increases significantly with A_L.* ***Column 2*** *denotes inventories where L/W increases slightly with A_L, and declines for the largest A_L.* ***Column 3*** *denotes the inventory where L/W declines with A_L.*

Figure 5.27 suggests that the relationship between L/W and landslide area differs between different 'types' of inventory; potentially related to the triggering mechanism, initiation

location on the slope and predominant type of landslide triggered. It is possible that L/W distribution in the Liguria inventory (**Figure 5.27, Column 2**) actually behaves more like an inventory in **Figure 5.27, Column 1**. It is somewhat difficult to objectively categorise the behaviour of L/W in the largest landslide area categories due to few observations of very large landslides (> 12,800 m^2) in all inventories. In the following **Section 5.6.4.3**, we show the inverse Gamma pdf fit to L/W in each landslide area category and investigate variability in parameter values and pdf characteristics for each inventory.

5.6.4.3 Results of Inverse Gamma Probability Density Function Fitting to Length-to-width Ratio in Landslide Area Categories for Each Additional Inventory

Figure 5.28 shows the results of Monte Carlo Kolmogorov-Smirnov goodness-of-fit testing (**Section 5.6.2.3**) for each landslide area category and each inventory. In all cases, the inverse Gamma fit can be considered reasonable, as the K-S statistic (D_{OBS}) lies below the 90th percentile of D_{SIM} (**Figure 5.28**). In **Figure 5.28**, we show results for all inventories and landslide area categories where there is data, but we use a transparency to denote landslide area categories where there are $n < 40$ observations of L/W. The K-S test is more 'lenient' when there are fewer observations and D_{SIM} values are generally higher and occupy a wider range. Thus although these results appear to show a statistically significant fit, we acknowledge that results in these categories will be less robust.

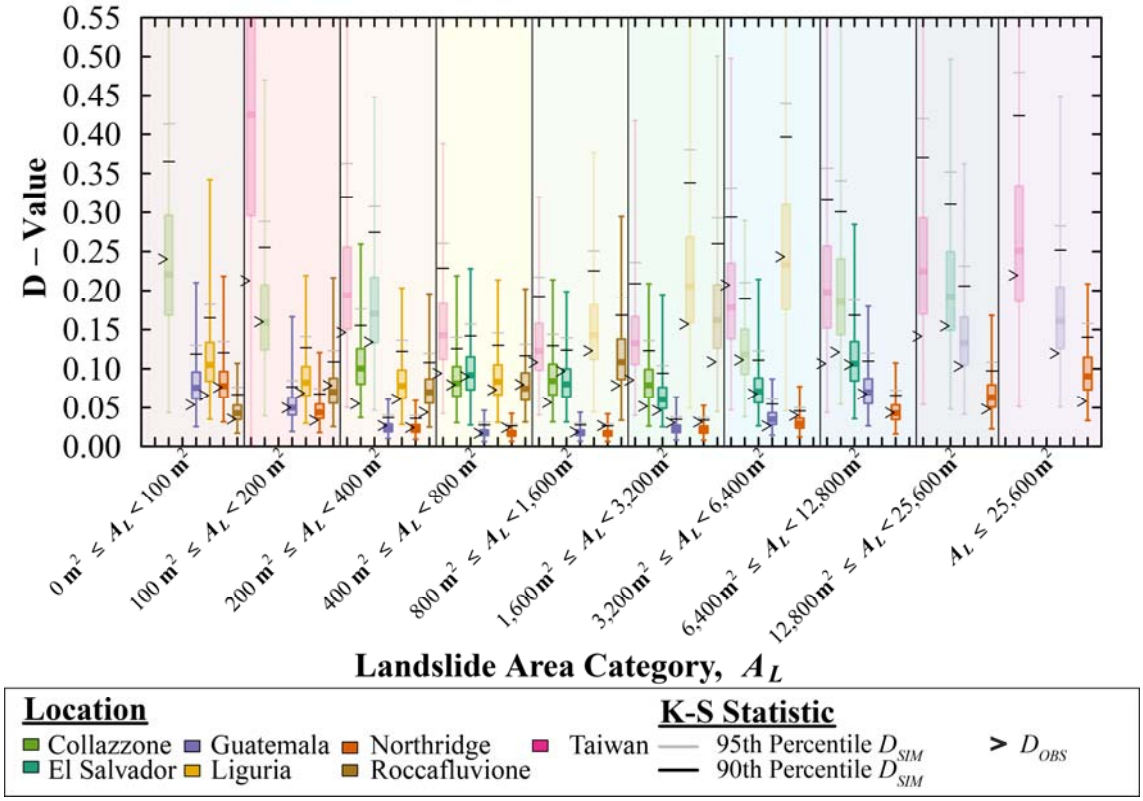


Figure 5.28 *Results of Monte Carlo Kolmogorov-Smirnov goodness-of-fit testing (**Section 5.6.2.3**) for each landslide area category and each inventory. D_{OBS} represents the maximum distance between the inverse Gamma cumulative distribution function and the observed data. D_{SIM} is the maximum distance between random deviates selected from an inverse Gamma pdf with known parameter values and an inverse Gamma pdf fit to these random deviates. D_{SIM} boxplots represent 2,500 iterations.*

Figure 5.29 shows each inverse Gamma pdf fit to L/W in each landslide category where there are > 40 observations of L/W. In the Taiwan inventory, no category has more than 40 observations, so only those with ≥ 8 observations are shown (in semi-transparency).

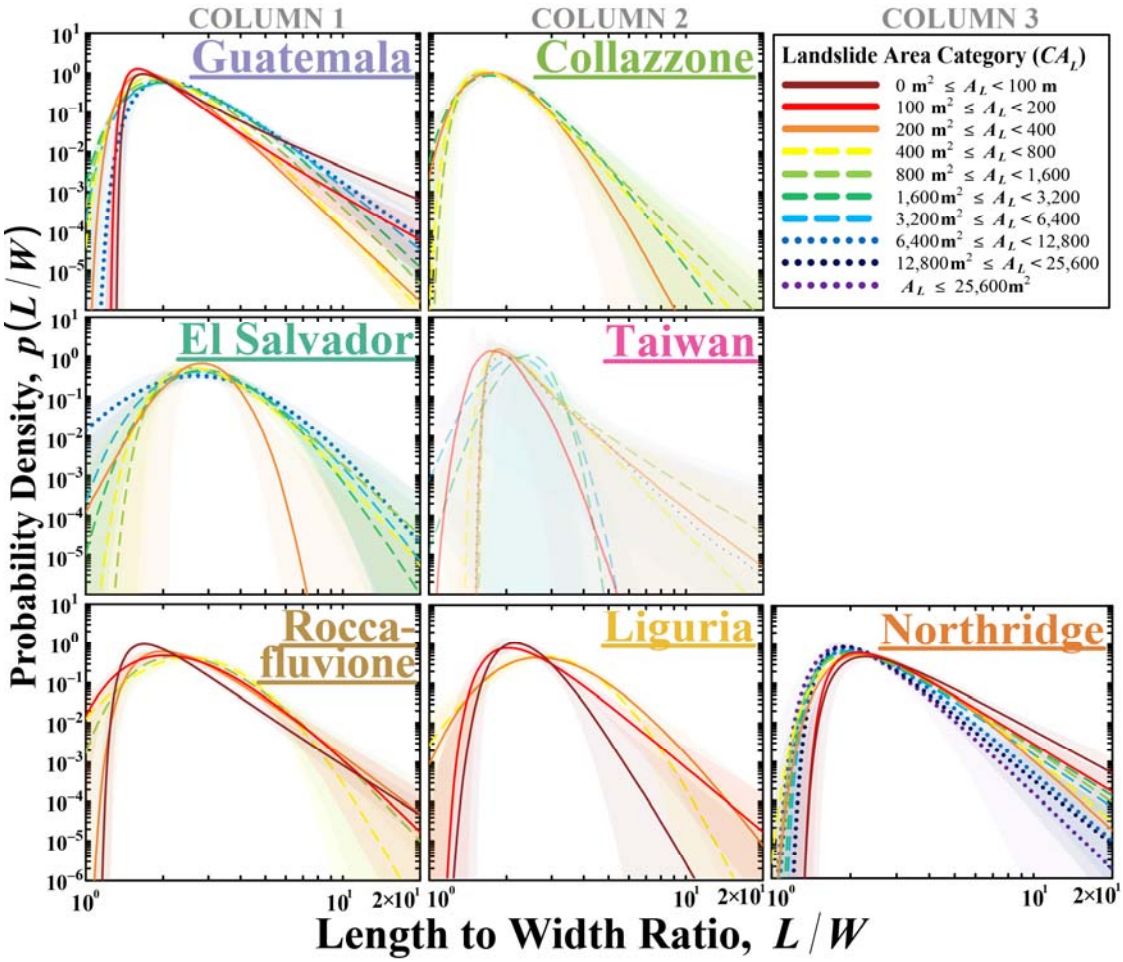
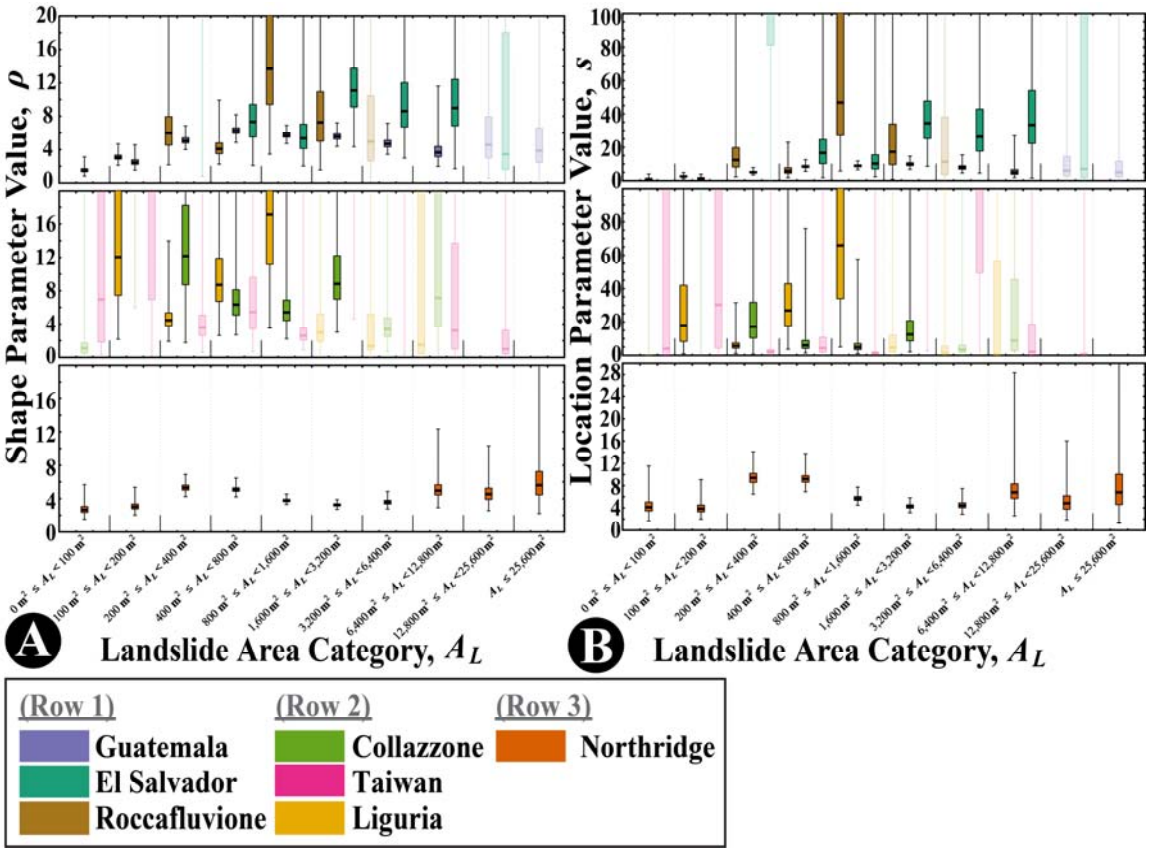


Figure 5.29 *Inverse Gamma pdfs fit to L/W distribution using a bootstrapped MLE technique (**Section 5.6.2.2**) in each landslide area category for each inventory. Inventories are ordered as in **Figure 5.27**, where **Column 1** denotes inventories where L/W increases significantly with A_L. **Column 2** denotes inventories where L/W increases slightly with A_L, and declines for the largest A_L. **Column 3** denotes the inventory where L/W declines with A_L. Only shown are pdfs fit on n > 40 observations of L/W. None of the area categories for Taiwan have n > 40 values, so these pdfs are shown semi-transparent.*

As in **Figure 5.27**, in **Figure 5.29** the inventories have been organised into three columns based on L/W behaviour. Broadly, inventories in **Column 1** tend to have more moderate right hand tail power law decays than inventories in **Column 2**, supporting the observation in **Figure 5.27** that these inventories have landslides with larger L/W values. Inventories in **Column 2** appear to have a steeper left hand tail than those in **Column 1**, suggesting there is relatively a lower probability of observing $L/W \approx 1.0$ (a compact, circular landslide) than inventories in **Column 1**. The shape of pdfs in Northridge inventory (**Column 3**) is visually quite similar to the Guatemala inventory, although these are inverse in terms of landslide category (L/W increases with A_L in Guatemala, L/W decreases with A_L in Northridge). In the following **Section 5.6.4.4** we investigate the parameter values describing each pdf.

5.6.4.4 Variability in inverse Gamma pdf parameter values

Assuming that the inverse Gamma pdf is a good model of L/W distribution within landslide area categories across all inventories (as demonstrated in **Figure 5.28**), we now investigate variability in inverse Gamma parameter values (ρ, a, s). **Figure 5.30** shows results of bootstrap MLE parameter estimation (**Section 5.6.2.2**) to fit inverse Gamma pdfs to each dataset. To aid visualisation of variability in parameters with landslide category, we have separated inventories into those with similar behaviour. As in **Figure 5.21**, we use a transparency to denote any landslide category from any inventory where there are $n < 40$ observations of L/W.



(*Figure 5.30 continues on the following page*)

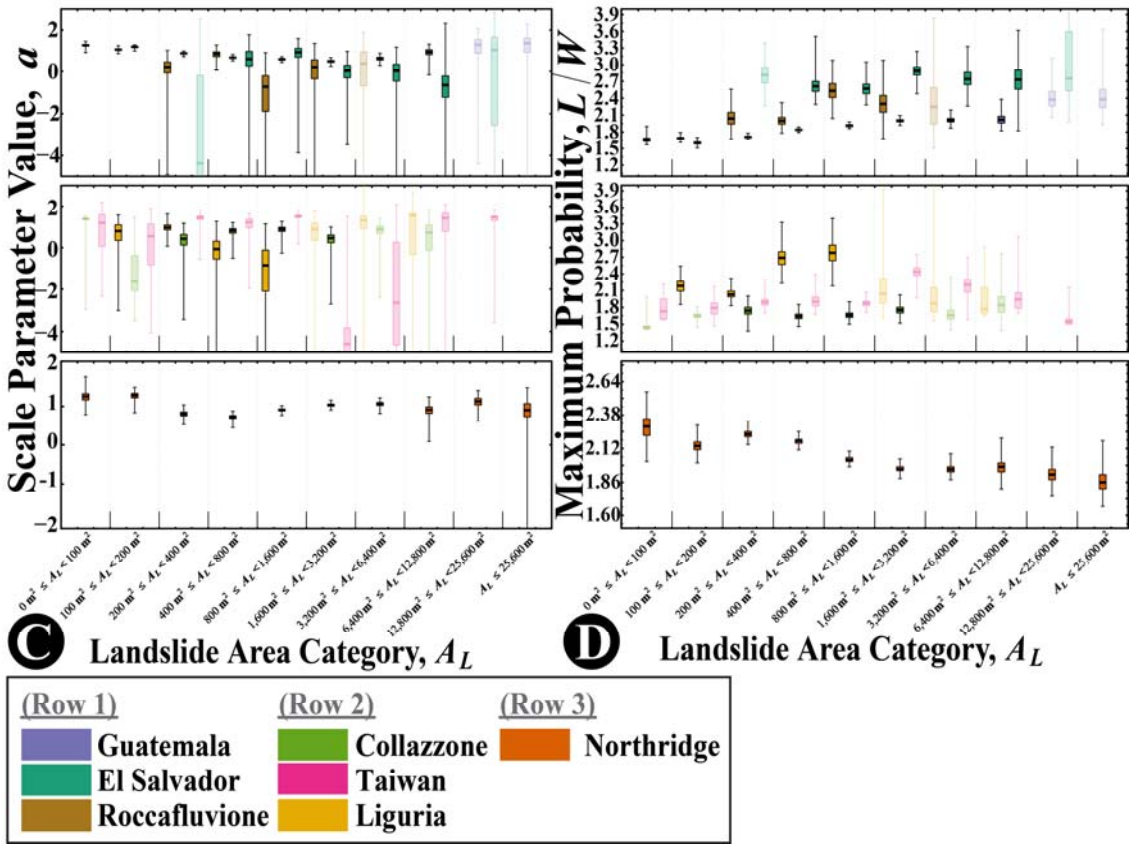


Figure 5.30 *Boxplots of shape (ρ), scale (a), location (s) parameter values and location of rollover (L/W) from bootstrapping MLE fitting of the inverse Gamma pdf to landslides in each area category for each inventory. To aid visualisation, inventories have been split into three categories, as in* ***Figure 5.22****, depending on relationship between landslide category A_L and L/W. Boxplots in semi-transparent colours denote parameter values where the inverse Gamma pdf has been fit to n < 40 observations of L/W.* ***(A)*** *Boxplots of Shape parameter (ρ) which primarily controls the inverse power law decay of the right hand tail.* ***(B)*** *Boxplots of scale (a) parameter value which primarily controls the L/W at which the maximum probability density occurs (the rollover)* ***(C)*** *Boxplots of location (s) parameter value which primarily controls the gradient of the left hand tail of the distribution.* ***(D)*** *Boxplots of the location of the rollover (the L/W at which the maximum probability occurs).*

Figure 5.30 demonstrates considerable variability in parameter values both between landslide area categories and between inventories, and it is difficult to discern whether there are general trends in *L/W* common to all inventories, or specific 'types' of inventory. Generally, larger numbers of observations result in smaller range of parameter values from bootstrapping (due to a larger sample size effectively being more like a population than a sample, and improved power of MLE fitting), as seen for the Northridge and Guatemala inventories. Indeed, a very wide range of a parameter value in an individual landslide area category (such as those observed for Roccafluvione and Liguria) suggests that the sample size is somewhat inappropriately small for this type of statistical analysis, rather than giving any insight into physical process. Indeed, we have cut the axes of **Figure 5.30** at less than the interquartile range for some boxplots, as these extended by a nonsensical amount.

The shape parameter (ρ) is generally higher for inventories triggered by moisture (rainfall or snowmelt: El Salvador, Roccafluvione, Collazzone, Taiwan and Liguria), meaning that the right hand inverse power law tail decays more rapidly (and thus there is a lower probability of observing large values of L/W) than in the Northridge inventory. From visual inspection of the inventories, it is possible that this is not actually physically based, but the result of a greater number of debris flow type landslides not being well approximated by an ellipse, and thus removed from analysis. Indeed, in the Liguria inventory we observe some of the largest ρ values, but in this inventory 34 % of landslides were removed because they were poorly approximated by an ellipse. Many of the rejected landslides were long and thin (and would thus have a high L/W value) but were also sinuous or branched, and thus poorly approximated by an ellipse. Consequently, it is possible that the ρ values are artificially high in inventories with a large proportion of debris flows that tend to follow drainage morphology (Dikau *et al.* 1996).

There does not appear to be a particularly discernible trend in location parameter (s) between the additional inventories (Collazzone, El Salvador, Liguria, Roccafluvione and Taiwan). Location primarily controls the left hand tail of the distribution; a high value denoting a gentle gradient and a higher probability of small values of L/W, a low value denoting a steep gradient and a lower probability of small values of L/W. Both the Collazzone and El Salvador inventories roughly follow the trend in s parameter variation with A_L that was observed for Northridge and Guatemala: decreasing with increasing A_L for landslide areas between 400 m^2 – 3,200 m^2 and then increasing with increasing A_L, although the actual s values are generally higher than observed in Northridge and Guatemala. Both the Roccafluvione and Liguria inventories have considerably higher s parameter values, particularly for $800 < A_L \leq 1{,}600$ m^2, suggesting a higher probability of compact landslide shapes for medium sized landslide areas.

Figure 5.30C shows the scale parameter (a), which primarily controls the location of the rollover (maximum probability L/W). Scale (a) shows a somewhat more discernible trend than the parameters that control the tails of the distribution (ρ, s). For the water-triggered inventories, scale generally declines with increasing A_L, although does tend to slightly increase for the very largest landslide area categories (potentially a result of few observations of very large landslides, and removal of large, long debris flows with low ellipticity values). This results in an increasing L/W at which the maximum probability (rollover) occurs with increasing A_L (**Figure 5.30D**), demonstrating that these inventories

tend towards larger values of L/W for larger landslide areas, and thus large landslides tend to be longer and thinner. For the Northridge earthquake triggered inventory, a occupies a narrower range and first declines and then increases with A_L. Overall, this results in a declining value of L/W rollover position with increasing landslide area. That is, the larger the landslide area, the more likely the shape is to be compact.

Overall this **Section 5.6.4.4** has shown that the inverse Gamma pdf does appear to be an appropriate statistical model of landslide L/W for each landslide category in each inventory. However, **Figure 5.30** demonstrates that the tails of the inverse Gamma pdf modelling landslide length-to-width ratio seem to vary between inventories and do not show a strong signature of a 'generally applicable' underlying pattern. There does appear to be a more discernible trend in increasing value of L/W at which the maximum probability occurs with increasing A_L for the rainfall-triggered inventories. In the following **Section 5.6.4.5** we briefly investigate whether the inverse Gamma pdf models the entire distribution of L/W for each landslide inventory.

5.6.4.5 Fitting one Inverse Gamma Probability Density Function to all values of Length-to-width Ratio for Each Inventory

In this **Section 5.6.4.5** we briefly assess whether one inverse Gamma probability density function (pdf) models the distribution of all values of length-to-width ratio (L/W) in each inventory (i.e., regardless of landslide area category). **Figure 5.31** shows the results of the Monte Carlo Kolmogorov-Smirnov goodness-of-fit testing and **Figure 5.32** shows the inverse Gamma pdf fit to each inventory.

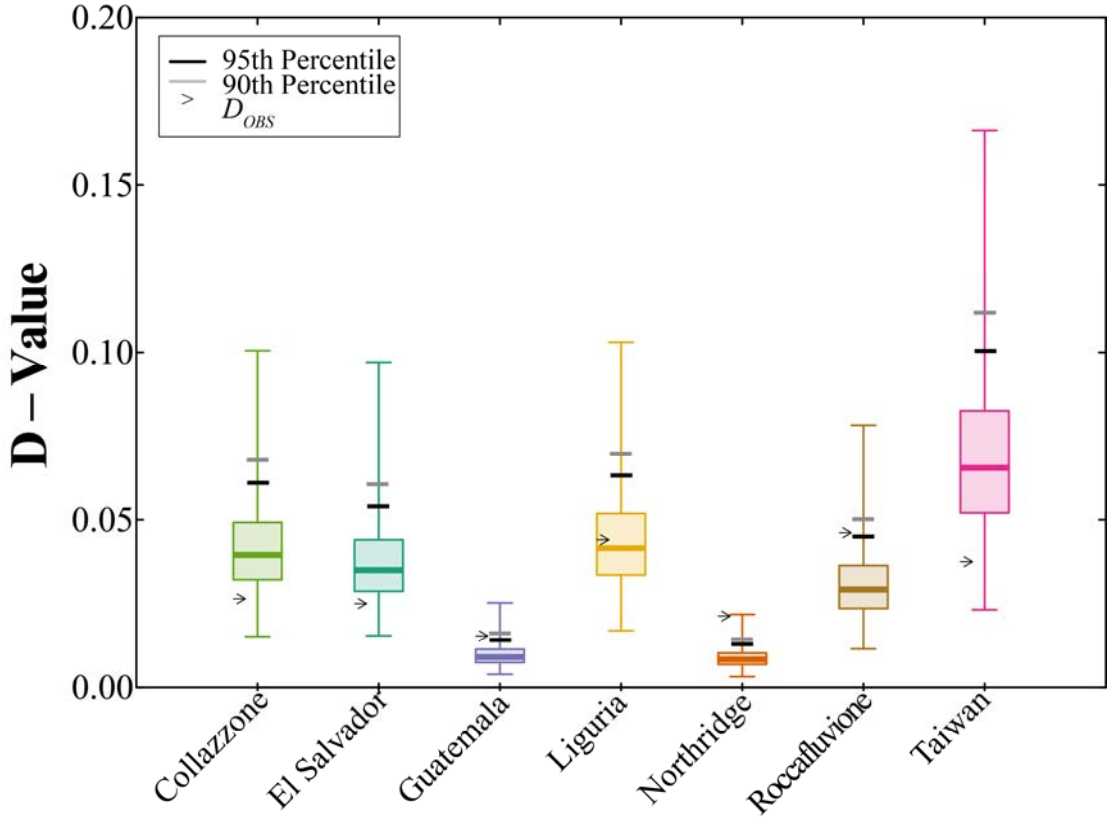


Figure 5.31 *Results of Monte Carlo Kolmogorov-Smirnov goodness-of-fit testing (**Section 5.5.1.3**) for each landslide inventory. D_{OBS} represents the maximum distance between the inverse Gamma cumulative distribution function and the observed data. D_{SIM} is the maximum distance between random deviates selected from an inverse Gamma pdf with known parameter values and an inverse Gamma pdf fit to these random deviates. D_{SIM} boxplots represent 2,500 iterations.*

Figure 5.31 shows that for all inventories other than the Northridge inventory, the inverse Gamma pdf is a reasonable model of the entire distribution of L/W in an inventory. We see that for the inventories with the largest numbers of observations (Guatemala and Northridge), the range of simulated D-values is much narrower, due to the improved power of MLE fitting when there are more observations. The fact that the distribution of L/W in the Northridge inventory is not well modelled by the inverse Gamma pdf could allude to some different physical process governing the shape of landslides in earthquake triggered inventories – as trends in L/W with landslide area category were also different compared to all other inventories investigated (**Section 5.6.3**).

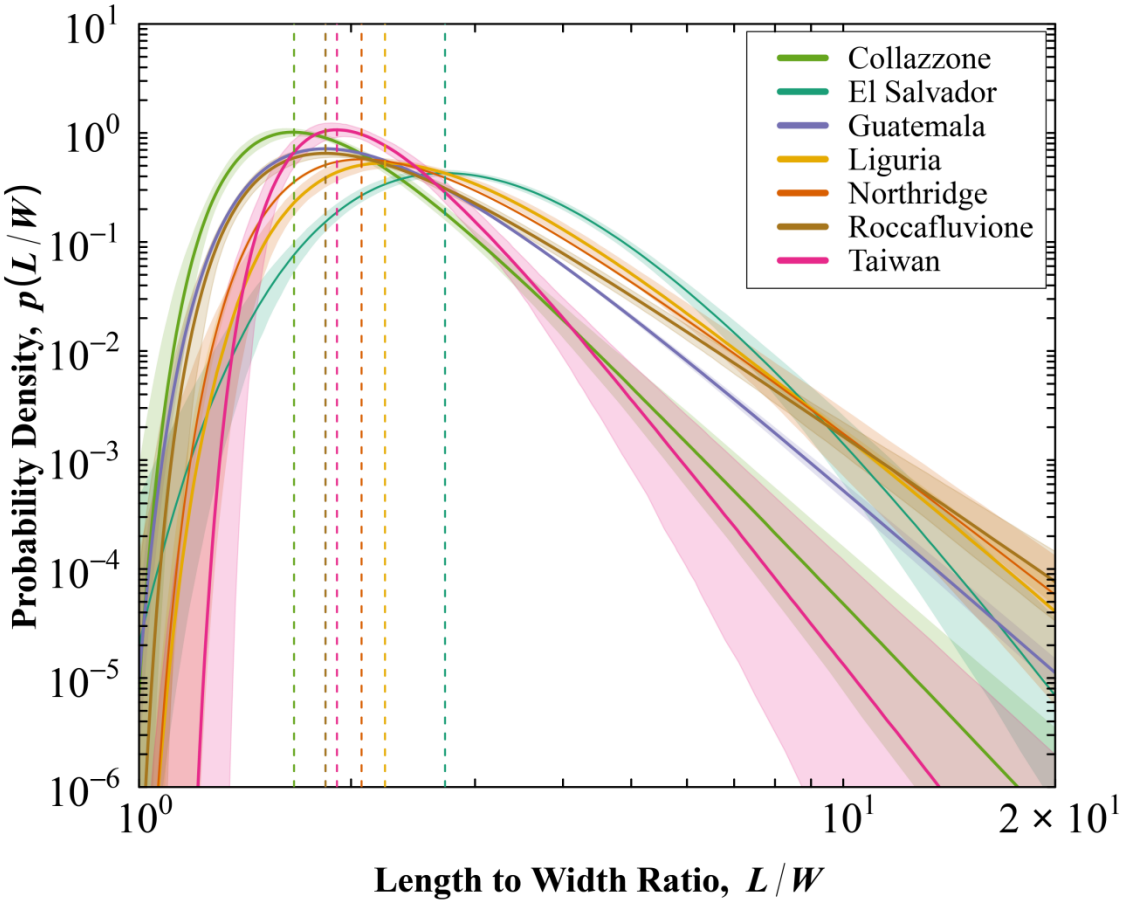


Figure 5.32 *The inverse Gamma probability density function (**Eq. 5.5**) fit to landslide ellipse L/W using maximum likelihood estimation (**Section 5.6.2.1**) landslide inventories listed in **Table 5.2**. The shaded area either side of the pdf line represents a form of 5th/95th percentile confidence intervals around the pdf. These are calculated using a bootstrapping technique (**Section 5.6.2.2**) where observed data is repeatedly sampled with replacement and an inverse Gamma pdf fit to each sample. Parameter values stated are medians from bootstrapping technique plus minus one standard deviation. The location of the median rollover (from bootstrapping technique) for each inventory is also shown as a dashed vertical line.*

Figure 5.32 shows some similar overall behaviour in L/W distribution for some of the inventories: the Liguria, Northridge, Guatemala and Roccafluvione pdfs follow a similar shape and the Taiwan and Collazzone pdfs are similar but with a steeper inverse power law decay tail for large values of L/W. There is quite a broad range of L/W values at which the rollover occurs – the smallest L/W is Collazzone, and the largest is El Salvador. These results are roughly in line with the findings in **Section 5.6.4.3** that the Collazzone and Taiwan inventories occupy a narrower range of L/W than the other inventories.

5.6.4.6 Section Conclusions

Overall, **Figure 5.30** demonstrates that the tails of the inverse Gamma pdf modelling landslide length-to-width ratio seem to vary between inventories and do not show a strong signature of a 'generally applicable' underlying pattern. We have touched upon the idea that

this could be attributable to methodological issues rather than physical process. Those issues are:

- *Completeness of the inventory* and under sampling of small landslide areas
- *Certain types of landslide* being preferentially rejected as they are not well modelled by an ellipse
- *Small numbers of observations,* particularly for very small and large landslide areas

There does appear to be a more general trend in the L/W at which the maximum probability density (the rollover) occurs: for water-triggered inventories, L/W increases with A_L, whereas for the one earthquake triggered inventory we have access to, L/W decreases with increasing A_L. This generally observed trend in rollover location could add weight to the argument that lack of coherent trends in the tails of the distribution is more a result of data quality and quantity rather than physical process, as the location of the rollover is determined by where the majority of data points lie, and this is where we observe a clearer trend across all inventories. Where we have fewer observations (in the tails of the distributions), there is less of a clear overall trend.

Analysis of additional large, substantially complete triggered event inventories could potentially decide whether there are generally observable patterns in landslide L/W distribution across a variety of inventories, or if L/W distribution genuinely does differ between inventories, perhaps as a result of physical processes. A second large, substantially complete earthquake triggered event inventory could compare whether the same trends are observed in the Northridge inventory. Although we have analysed several water-triggered event inventories, none of these can be considered as 'rigorous' as the Guatemala Hurricane Mitch Inventory, so analysis would certainly benefit from additional high confidence inventories. Moreover, a method to systematically and rapidly measure the L/W of sinuous, non-elliptical debris flows is required. This is not performed in this thesis, as the analysis performed in **Chapter 5** was motivated by the need to have a standard shape to model landslides as in the Landslide Road Impact Model (LRIM) (**Chapter 7**). However, this is certainly something we would like to investigate further.

In **Section 5.7**, we begin to discuss some of the potential physical reasons behind the trends observed in landslide L/W and compare our results to those found in current literature.

5.7 Discussion

5.7.1 Summary of Findings

Section 5.5 outlined a method to quantify the shape of landslides using a single index (ellipse length-to-width ratio, L/W) which can be rapidly applied to all landslides in a triggered event inventory in a standardised way. **Section 5.6** showed the results and analysis of applying this method first to two substantially complete, large triggered landslide event inventories which we consider to be 'high confidence' in terms of data quality and sample size for robust statistical analysis. We found that there was a statistically significant difference in L/W distribution when landslides are separated into categories depending on their area (A_L). Using maximum likelihood estimation and a Monte-Carlo Kolmogorov-Smirnov goodness-of-fit testing, we found that an inverse Gamma probability density function models well the range and distribution of L/W within each area category, and the parameters of this pdf vary with A_L. Although the absolute parameter values differ, the overall trend in variation of parameter values was quite similar between the Northridge and Guatemala inventories. The overall behaviour of L/W with A_L differs between the inventories; in Guatemala, as landslides get bigger, they tend towards longer and thinner shapes, whereas in Northridge, larger landslides tend to be more compact.

We then showed the application of the method to five additional triggered landslide event inventories (**Table 5.2**) which are considered 'lower confidence' either because methods of inventory production and/or small sample sizes creating some uncertainty in statistical analysis of results. The overall behaviour of L/W with A_L appears to be similar for water triggered landslide inventories, although the actual parameter values vary, and there is uncertainty due to smaller sample sizes. We only have one earthquake triggered inventory (Northridge), so further testing is required to see if the same trends are observed in other inventories. Ideally, we would like to apply the analysis to further large, high confidence inventories to verify whether there are generally observable trends in landslide shape. In the following **Section 5.7.2** we critically discuss findings and offer some potential physical reasons for the behaviour we observe.

5.7.2 Discussion of Findings and Potential Reasons for Observed Landslide Length-to-width Behaviour

5.7.2.1 Modelling Landslides as Ellipses

For each inventory (listed in **Table 5.2**), we found that an ellipse was a reasonably good approximation of landslide shape for 66–96 % of landslide polygons in each inventory, when using a threshold value of ellipticity goodness-of-fit of $e \geq 0.5$ (**Eq. 5.4**). We believe the majority of landslides whose shape is not well approximated by an ellipse are long, sinuous debris flow run out zones which follow channel morphology. Indeed, in inventories where the majority of landslides can be considered 'low mobility' (i.e., more dominated by rotational and translational slides with less run out), the proportion of landslides that were well modelled by an ellipse is generally > 85 %. The removal of landslides not well modelled by an ellipse introduces a bias by generally removing long, thin landslides that would typically have high values of L/W. However, Klar *et al.* (2011) suggest that maps of the landslide scar and landslide deposit represent different processes, the latter being additionally affected by setting, topography, velocity, etc. Although our method does not distinguish between the scar and the deposits in a deterministic way, it can be considered somewhat of a proxy.

The method of measuring L/W from ellipses assumes that all landslides are longer than they are wide, and thus length-to-width ratio must be $L/W \geq 1.0$. Although there are a small number of cases where this is unrealistic, we consider this is a reasonable simplification for the sake of reducing processing time and data requirements, as several authors find that the majority of landslides are longer than they are wide (Gabet and Dunne, 2002; Rickli *et al.*, 2008; Marchesini *et al.*, 2009; Milledge *et al.*, 2014).

From the literature (**Table 5.1**), we found that landslides were often described as 'elliptical' but this had not been quantitatively tested. Moreover, in those studies that quantified landslide length-to-width ratio, it was not always clear how this was being measured, creating difficulty in comparing results of different authors. The methods and results in this **Chapter 5** present a more 'statistically robust' confirmation that the majority of landslide shapes are roughly elliptical, so this is an appropriate shape for landslides that will be introduced into the Landslide Road Impact Model (LRIM) in **Chapter 7**. More broadly, we have developed an automated and standardised technique to quantify landslide shape, which can now be applied to other inventories to compare results.

5.7.2.2 Observed Behaviour of Landslide Length-to-width Ratio Compared to Existing Literature

Section 5.6 showed median values of L/W range 1.9 – 3.2 depending on the inventory, which is line with typical values of L/W stated in the literature reviewed in **Table 5.1**, where median values stated were 1.5 (Rickli and Graf, 2009) and average L/W values range 0.57 – 2.8 (Reneau *et al.* 1987; Cardinali *et al.* 1999; Parise and Jibson, 2000; Gabet and Dunne, 2002; Süzen, 2002; Rickli and Graf, 2009; Quinn *et al.* 2011). Where the research presented in this chapter does deviate from the existing literature is identifying the presence of larger values of L/W. Maximum values of L/W identified in the literature typically in the range $4 < L/W < 8$ (Moine *et al.* 2009; Pourghasemi *et al.,* 2013; Yang and Lee, 2006). Whereas, we observed landslides where $L/W > 8$ in all but two of the inventories analysed (Collazzone and Taiwan). This could potentially be attributed to the inclusion of both source area and run-out in our analyses. Typically landslides with $L/W > 8$ represent < 1.5 % of each inventory, but the heavy tailed decay of large values of L/W has important implications for risk forecasting and modelling, as it is likely that these very long run out landslides have the potential to be destructive and fast moving (Legros, 2002). Moreover, it is likely that the inverse power law decay heavy tail would actually decay more slowly than shown here, due to the removal of landslides that were not modelled well by an ellipse. Indeed, in experiments measuring the L/W of a small sample of rejected landslides by hand, we identified landslides where $L/W > 50$.

The findings here also depart from some broadly accepted 'rules of thumb' about landslide shape being roughly scale invariant. For example, Hovius *et al.* (1997) estimate that landslide width is roughly proportional to the square root of its area, Guthrie *et al.* (2008) approximate that landslide length is proportional to $0.76 \times A_L^{0.66}$ and Frattini and Crosta (2013) state that landslide width is generally equal to $L/2$. These typical estimates of landslide dimensions are widely used for modelling and remote sensing purposes. By fitting an inverse Gamma probability density function to L/W separated into categories of landslide area, we have shown that the parameter values and shape of the distribution of L/W varies with area, and thus we can say that landslide shape varies with landslide area. This supports the findings of Klar *et al.* (2011) who proposed that landslides have 'self-affine' geometry where the depth of a landslide generally decreases non-linearly with increasing landslide length.

5.7.2.3 Potential Physical Explanations for Observed Behaviour of Landslide Length-to-width Ratio

Regarding the physical processes resulting in landslides of different shape, authors have suggested that the following factors will affect either the length or width (or both) of landslides: root cohesion (Casadei *et al.,* 2003); wet or dry triggering mechanism (Legros, 2002; Yang and Lee 2006); topography (Guthrie and Evans, 2004); soil characteristics (Cardinali *et al.,* 2000; Klar *et al.,* 2011; Lehman and Or 2012; Milledge *et al.,* 2014). Milledge *et al.* (2014) also showed the probability density distribution of landslide L/W for six different landslide inventories (a mix of triggered event, seasonal and historical inventories) varied quite considerably in terms of skew and kurtosis. In contrast, other authors have found that landslide dimensions do not significantly vary with factors such as vegetation cover (Reneau *et al.,* 1987; Rickli and Graf, 2009) and presence of water (Martel, 2004).

Broadly, we found that the L/W at which the maximum probability occurs (the rollover) tends to increase with landslide area for all inventories we investigated that were triggered by moisture. This is supported by Milledge *et al.* (2014)'s recent work, which builds upon Dietrich *et al.* (2008) to create a multidimensional slope stability model for shallow landslides and investigate the relationship between landslide shape and soil strength. Milledge *et al.* (2014) found that the minimum area required for a slope to fail increased with landslide L/W. They also found that landslide width is one of the principal controlling factors of whether a landslide of a given size will initiate – as the blocks above and below the landslide parcel mainly control resistance to failure, and the depth of this block is mainly controlled by soil depth. Thus landslide width determines the surface area in contact with these up and down slope land parcels. Milledge *et al.* (2014) found that the least stable landslide shape (in terms of L/W) increases with failure plane depth, as a smaller landslide width results in less strength from the contact between the toe end of the landslide and the downslope land parcel. Ultimately, Milledge *et al.* (2014) state that the shape of an unstable shallow parcel of land will be determined by the spatial pattern of soil strength and water content across a landscape. In **Chapter 4**, we discussed Pelletier *et al.* (1997)'s findings that patches of soil with a moisture content greater than a threshold value followed an inverse-power law decay which may exert some control over 'generally observed' landslide frequency-area statistics. Combined with Milledge *et al.* (2014)'s findings about the relationship between soil strength, moisture and landslide shape, this could potentially point towards a relationship between the 'generally observed' landslide area frequency size

statistics and the findings of this **Chapter 5**, which would certainly warrant further investigation.

Regarding the variability in probability density functions between the different inventories, we cannot confidently state whether this is a result of process and setting or simply an issue of data quantity and quality. Potentially, the length of landslides could be controlled by the length of the slope (Klar *et al.*, 2011) and the location on the slope at which the landslides initiate (Guthrie and Evans, 2004). Although we have not calculated precisely where on the slope the landslides typically are located (as per Meunier *et al.*, 2008), some information is included in the inventories (**Table 5.2**), which could support this theory:

- *Landslides in El Salvador* typically travelled from top to base of the slope (Crone *et al.* 2001). This could explain large values of L/W for large landslide areas in this location.
- *Landslides in Liguria* were typically located at the foot of slopes (Mondini, 2014) which would suggest that long run-out landslides with high values of L/W were either (i) prevented by intersection with drainage channels or (ii) diverted into drainage channels, creating sinuous, asymmetric landslide polygons that were removed from analysis.
- *Landslides in Collazzone* were prevented from mobilising into debris flows because of the high clay content of the soil (Cardinali *et al.*, 2000), potentially explaining the relatively low and narrow range of L/W values in this inventory.
- *Landslides in Taiwan* typically occurred in the upper portion of slopes (Chen, 2014, personal communication) and were high mobility. This does not fit well with the observation of a low and narrow range of L/W values at this location.

Clearly, more work is required to better understand the physical controls over landslide shape This could involve the analysis of parameters such as soil strength and topography for the inventories already investigated here, or the analysis of additional large, substantially complete triggered event landslide inventories. We have highlighted potential linkages between the potential physical controls over observed 'general' behaviour of landslide frequency-area statistics and the physical controls over our observed trends in landslide length-to-width ratio. However, at present, the data we have used and the results we have found are not statistically robust enough to suggest that landslide shape displays fully 'generally applicable' behaviour, but is perhaps semi 'generally applicable' when inventories are separated by triggering mechanism. This is something we intend to investigate further.

5.7.2.4 Future Applications of the Method

In terms of the approach taken in this **Chapter 5**, we believe this is a relatively novel piece of research both in terms of the methods used and the amount of data analysed. Certainly, many papers give some summary statistics of L/W in the inventories they use (e.g., Barlow *et al.,* 2006; Parise and Jibson, 2000; Pourghasemi *et al.*, 2014), and a small number include tables of the dimensions of all landslides in an inventory, but this is often more a description of the data than a major piece of research in itself. Moreover, it is not always explained how the length and width dimensions of landslides have been calculated (e.g., Milledge *et al.,* 2014). As we have shown in **Section 5.5**, there are many ways to measure the length-to-width ratio of a landslide and these can give different results. The workflow we have developed in **Sections 5.5** and **5.6** presents a systematic and repeatable way to measure and analyse trends in L/W and compare 'like for like' between inventories. We have created scripts to automate this analysis in open-source software which can now be made publically available. The calculation of best fit landslide ellipse and their L/W is a Python code which runs in GRASS-GIS, and can now be uploaded to the GRASS-GIS Add-ons Repository. The statistical analysis described in **Section 5.6** to fit an inverse Gamma pdf to L/W data and produce visualisations is a series of R scripts which can be distributed online, or potentially incorporated into existing Web Processing Services that analyse the frequency-size statistics of landslide area (Rossi *et al.,* 2012; Rossi and Malamud, 2014). By sharing the tools and methodology to analyse L/W in this way, we hope that others may apply this approach to their own 'high confidence' landslide inventories and either the similarities or differences in L/W behaviour between inventories may become clearer.

5.7.3 Implications for Landslide Road Impact Model

The original motivation for performing this research was the question of what shape landslides should be in the Landslide Road Impact Model (LRIM) (**Chapter 7**). This shape must be easy to model in a GIS (i.e., a regular shape that does make large demands on processing power by e.g., querying a DEM) and preferably reflect the typical shape of triggered landslides across a range of settings, thus allowing LRIM to be applied to various locations with minimal 'tuning' of the model. From literature reviewed and analysis of the inventories in **Section 5.5**, we are satisfied that an ellipse is a reasonable approximation of the shape of most landslides in a triggered landslide event inventory. **Section 5.6** showed that the L/W of a landslide ellipse can vary $1.2 < L/W < 15.1$ and there appears to be a statistically significant variation in L/W with landslide area, which is well modelled by a

three-parameter inverse Gamma pdf. In LRIM, we can now instigate the following protocol:

1. *Generate a landslide area* (A_L) by generating a random deviate from the inverse Gamma pdf already established for landslide areas (Malamud *et al.*, 2004a).
2. *Select the inverse Gamma pdf that models L/W distribution* for the landslide area category that the landslide area (A_L) belongs to.
3. *Generate a random value of L/W* from this inverse Gamma pdf.
4. *Calculate the length of the long and short axes* of this ellipse from L/W and A_L.
5. *Plot this ellipse and repeat Steps 1 – 4* for each landslide to be modelled.

Step 2 is somewhat problematic as we did not identify one set of parameters that modelled L/W distribution within a landslide area category across all inventories (i.e., we do not observe 'generally applicable' behaviour in the relationship between landslide shape and size). At present, in LRIM, we select the inverse Gamma parameters that model L/W for the particular location we are investigating (i.e., we use the Northridge pdfs to apply LRIM to Northridge), but this would not be possible when applying LRIM to locations where a triggered event inventory is not available. Potentially, future users could select parameters based on the inventory that best reflects the environment they are modelling (e.g., in a region susceptible to debris flows, the Guatemala parameter values may be most suitable), or calculate and use their own parameter values by using the GRASS-GIS add-on and R script to analyse landslide L/W if any inventories are available. We hope that through analysis of more inventories in the future, we may be able to either (i) decide on a single set of parameter values for each landslide area category that model L/W for all locations or (ii) allow the user to select from a set of descriptive categories a set of parameter values that reflect L/W distribution for the inventory they wish to model.

5.8 Conclusions and Chapter 5 Summary

In this Chapter 5 we:

5.2 Discussed how shape can be measured quantitatively across different disciplines.
- We found that indices based on shape dimensions (e.g., perimeter, area, length, width) are commonly used and are a simple, easily applicable way to simplify a complex shape to a number which can be compared with other shapes.

5.3 Reviewed the literature discussing what shape landslides typically are.

- We found that commonly, landslide shape is often described qualitatively as approximately elliptical.
- Where landslide shape is investigated numerically, this is typically measured from the ratio of length-to-width.
- In the literature, landslide length-to-width ratio ranges $0.28 < L/W < 8.0$ and is generally $L/W < 5.0$.

5.4 Introduced the landslide inventories used in analysis:

- We first described two large, substantially complete triggered landslide event inventories: 11,111 landslides triggered by the 1994 Northridge Earthquake, USA and 9,594 landslides triggered by 1998 Hurricane Mitch in Guatemala.
- We assume these inventories represent almost the entire 'population' of landslide shapes in a triggered event, and can be used to establish a methodology to measure and statistically analyse landslide shape.
- We then introduce five additional inventories (413 landslides triggered by 1997 snowmelt in Collazzone, Italy; 621 landslides triggered by the 1998 Hurricane Mitch in El Salvador; 537 landslides triggered by heavy rainfall in 2011 in Liguria, Italy; 851 landslides triggered by heavy rainfall and snowmelt in 2013 in Roccafluvione, Italy and 140 landslides triggered by the 2010 Typhoon Megi in Shu-Wa, Taiwan), which are an order of magnitude smaller in terms of number of observations, and some can be considered lower confidence in terms of completeness and methodology used.

5.5 Experimented with different methods of abstracting the shape of landslides in large inventories to ellipses and testing the goodness-of-fit of that ellipse.

- We found that a method that effectively represents many landslide polygons is by fitting a convex hull to a landslide polygon, creating an ellipse based on the area and perimeter of that convex hull and then scaling the ellipse to match the area of the original landslide.
- Generally, this method works well for > 80 % of landslide polygons in an inventory.

5.6 Investigated how landslide length-to-width varies as a function of landslide area.

- We found that there is a statistically significant difference in L/W distribution when landslides are separated into categories of area, using approximately logarithmically increasing bin sizes.
- We find that an inverse Gamma probability density function well models L/W distribution.
- We developed methods using Maximum Likelihood Estimation to fit an inverse Gamma probability density function to the distribution of L/W, test the goodness-of-fit of this distribution using a Monte-Carlo Kolmogorov-Smirnov test, and a Bootstrapping technique to estimate the uncertainty in probability density function parameter values.
- This method was initially applied to the Northridge and Guatemala Inventories, where we found that the trend in inverse Gamma parameter values is approximately similar between the inventories with increasing

landslide area. Although, overall, relationship between rollover location and landslide area category is inverse between the two inventories: in Northridge, the L/W at which the rollover occurs decreases with landslide area category, whereas in Guatemala, the rollover location increases with landslide area category.

- We then applied the analysis to five additional smaller, lower confidence inventories. Trends in inverse Gamma pdf fit to each landslide area category are less clear. We split inventories into three 'types' depending on L/W behaviour: inventories where L/W clearly increases with landslide area category, inventories where L/W slightly increases with landslide area category and inventories where L/W decreases with landslide area category. Inverse Gamma parameter values and location of rollover do show similarities within each inventory 'type'
- We believe these 'types' of L/W behaviour within different inventories could allude to different physical processes governing the shape of landslides in different locations. Although, it is possible that due to the small sample sizes and lower confidence in inventory production methods, these datasets may be 'noisy'. Certainly, this analysis must be applied to additional large, substantially complete and well documented triggered event inventories to test this.

5.7 Critically discussed potential physical reasons for observed L/W behaviour.

- We found that the range in L/W identified in this research is roughly in line with values of L/W quoted in the literature, although we observe some larger values of $L/W > 10$.
- The finding that L/W distribution varies with landslide area category appears to be relatively novel. It has generally been accepted that landslide shape scales regularly with area, but we have found that landslide L/W tends to increase with area for rainfall triggered inventories and decrease with area in earthquake triggered inventories.
- Our findings *of* L/W increasing with A_L appear to be in agreement with recent work by Milledge *et al.* (2014) looking at the physical relationship between landslide shape and stability. Milledge *et al.* (2014) found that the minimum landslide area required for a failure increased with increasing L/W and that landslide width plays a large factor in stability. This is due to landslide width and depth being the primary controls on the surface area (and resulting friction) of a sliding block. These potential physical explanations of landslide depth and friction suggest some parallels with the physical controls of generally applicable statistical distributions of landslide area (e.g., Stark and Guzzetti, 2009).
- We can now use the probability density functions established in this **Chapter 5** to model landslides as ellipses with variable values of L/W. In the LRIM model (**Chapter 7**), we randomly select a landslide area value from an already established statistical distribution. We then randomly generate a value of

L/W from the inverse Gamma pdf that corresponds to that landslide area category.

Chapter 6: Spatial Clustering in Triggered Landslide Event Inventories

Chapter Summary

In this **Chapter 6** we present two methods to quantify the spatial clustering of landslides in triggered landslide event inventories. These are (i) pair-correlation, where we measure the distances between all possible pairs of landslide centroids in a landslide inventory and (ii) the Kappa-measure which we develop here and is based on the normalized number of landslides per km^2 of a grid cell, applied in a moving window type approach. We apply pair-correlation to eight triggered landslide event inventories (413 landslides triggered by 1997 snowmelt in Collazzone, Italy; 9,594 landslides triggered by the 1998 Hurricane Mitch in Guatemala; 6,402 landslides triggered by the 1974 Guatemala City Earthquake, Guatemala; 537 landslides triggered by 2011 heavy rainfall in Liguria, Italy; 1,670 landslides triggered by 2009 heavy rainfall in Messina, Italy, 11,111 landslides triggered by the 1994 Northridge Earthquake, USA; 1,466 landslides triggered by the 2008 Typhoon Morakot in Taiwan; 4,222 landslides by 1997 snowmelt in Umbria, Italy) varying in total number of landslides, triggering mechanism, geographic setting and 'confidence' in the quality of the inventory. We find that for all eight inventories, the spatial pattern of landslide centroids measured using pair-correlation differs from a homogeneous distribution of randomly generated points, thus exhibiting spatial clustering. We find that when normalised by the extent of each respective study region, there are similarities in the overall pair-correlation distributions across all inventories. However, there appear to be slight differences between inventories dominated by 'low mobility' (slides and slips) and 'high mobility' (flows and falls) landslides, with low mobility triggered landslide event inventories clustering at a finer scale. We apply the Kappa-measure (normalized number of landslides per km^2) to two inventories where we also have maps of landslide susceptibility, and investigate the link between landslide clustering and landslide susceptibility. We find that clustering of landslides appears to occur at a finer spatial scale than modelled by most landslide susceptibility maps, and the relationship between landslide susceptibility of a zone and number of landslides that occur in that zone is non-linear and noisy. We show a simple method of incorporating the observed spatial patterns of clustering into the Landslide Road Impact Model (introduced in **Chapter 7**) to create synthetic triggered landslide event inventory maps, finding that incorporating clustering based on the results of pair-correlation does not significantly improve model skill in forecasting the number of road blockages resulting from a triggered landslide event.

List of abbreviations used in ***Chapter 6***

Abbreviation	Meaning	Section Introduced
LRIM	Landslide Road Impact Model	6.1

List of variables used in ***Chapter 6***

Variable	Meaning	Section Introduced
$\varkappa_{BL}$	Normalised number of road blockages (BL) per km^2	6.2
$\varkappa_{LS}$	Normalised number of landslide centroids per km^2	6.2
A_C	Area of a grid cell (C)	6.2
A_{Circ}	Area of a circular (Circ) study region	6.2
A_L	Individual landslide area	
A_S	Area of study region extent	6.2
L_D	Average landslide density across the study region	6.4
n	Number or count (different variable types)	6.4
(N, E)	Northing and Easting coordinates	6.1
N_{BL}	Number of road blockages (BL)	6.5
N_{Cells}	Total number of 1 km × 1 km grid cells required to cover the study region	6.2
N_{LS}	Total number of landslides (LS)	6.2
$p(x)$	Probability Density of x	6.4
r	Linear distance between pairs of centroids	6.2
S_{LS}	Landslide (LS) susceptibility	6.1
x	Random uniform value	6.1

6.1 Introduction

This **Chapter 6** presents a short summary of primary research into the spatial clustering of landslides in triggered landslide event inventories. This research was performed whilst developing the landslide road impact model (LRIM), which will be described fully in **Chapter 7**. The LRIM creates synthetic triggered landslide event inventories by randomly selecting landslide sizes and shapes from statistical probability density functions described in **Chapters 4** and **5** of this thesis. The way each synthetic landslide is given a location is discussed more fully in **Chapter 7**, but can be summarized as:

(i) *Randomly generating many thousands of pairs of (N, E) (where N = Northing and E = Easting) coordinates* from uniform distributions corresponding to the maximum N and E extent of the study region.

(ii) *For each random (N, E) coordinate, a random uniform value* $0 \leq x < 1$ is also generated.

(iii) *For each (N, E) point location generated, the value of landslide susceptibility ($0 \leq S_{LS} < 1$) is queried* at that location and compared to the random uniform value (x) generated in Step (ii).

(iv) *For each point location generated, if the random uniform value is greater than the landslide susceptibility, $x > S_{LS}$, the point is rejected.* If $x < S_{LS}$, the point is accepted and mapped.

(v) *The final number of desired synthetic landslide point locations is created from a random selection* of those accepted points from Step (iv).

From visual comparison (an example is shown in **Figure 6.1**) of LRIM model output maps of final points and observed landslide centroids from triggered landslide event inventories for the same locations, there appears to be a disparity in the spatial scale of clustering.

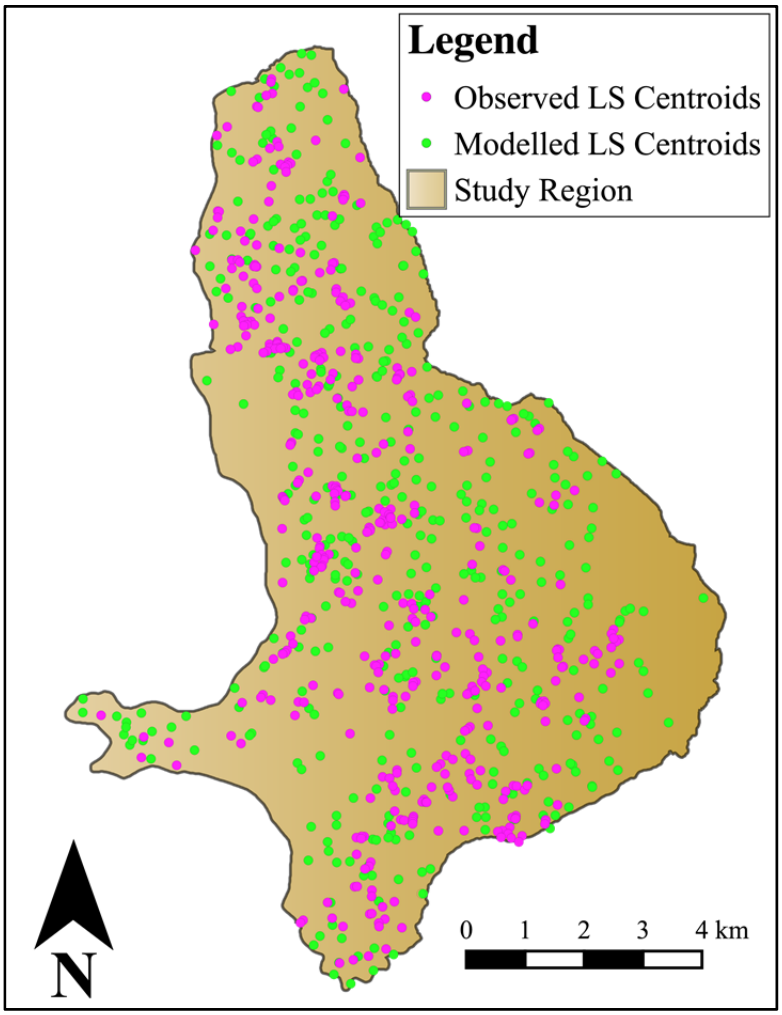


Figure 6.1 *Visual example of landslide spatial distribution for observed landslides in a triggered landslide event inventory of N_{LS} = 413 landslides triggered by 1997 snowmelt in Collazzone, Central Italy (pink dots) and N_{LS} = 413 modelled landslides from a synthetic triggered landslide event inventory generated by the Landslide Road Impact Model (LRIM) for the Collazzone region (described in* ***Chapter 7****) (green dots). The spatial clustering of observed landslides visually appears to occur at a finer scale than modelled landslides (i.e., modelled landslides are more spatially homogeneously distributed).*

The concept that landslides spatially cluster has been has been shown by others. For example, Cardinali (2000) and Guzzetti (2006a) both found that 75% of triggered landslides occur within 150 m of historical landslide locations in Central Italy. Zaitchik *et al.* (2003) found similar patterns for hurricane triggered landslides in Honduras, and found that using a landslide susceptibility map to forecast landslide locations did not generate the same scale of spatial clustering as observed in reality. Others that have noted spatial clustering in landslide inventories include Trauth *et al.* (2003) for Argentina; Jarman (2006) for Scotland and Tonini *et al.* (2013) for Switzerland.

In terms of physical explanations for observed patterns of clustering, authors such as Meunier *et al.* (2008) have shown that for earthquake triggered landslides, landslide clustering is linked to the pattern of ground motion and amplification from the earthquake. Meunier *et al.* (2008) also showed through the analysis of six different triggered landslide event inventories that different landslide triggers result in landslides clustering at different points on a slope. Similarly, Guthrie and Evans (2004) used an inventory of 201 debris flows occurring across a 2,861 km^2 region to show that landslides cluster in both space and time, most likely as a result of the movement of high-intensity storm cells across the region. Das (2011) measured the clustering of 178 landslide centroids occurring 1982–2009 across an 8 km^2 road corridor in Northern India and linked the spatial pattern of landslides to interactions between lithology, land cover, distance to roads, drainage density parameters.

To examine whether there may be a 'generally applicable' pattern of landslide clustering that could be used in the LRIM and whether this would significantly improve model skill in forecast road network impact, in this **Chapter 6** we:

(i) *Develop ways to quantify spatial clustering* of landslide centroids.
(ii) *Compare the spatial clustering* of landslide centroids in eight different triggered landslide event inventories.
(iii) *Compare synthetically generated landslide point maps* to real landslide inventory centroids.

Through comparison of real triggered landslide event inventories and LRIM model output, we show in **Section 6.4** that there does appear to be a finer scale of spatial clustering in reality than is modelled by using the LRIM steps outlined above. In **Section 6.5** through some initial experiments to incorporate landslide clustering into the LRIM, we show that

incorporating spatial clustering does not significantly improve LRIM output in terms of forecasting number of road blockages caused by model synthetic landslides when confronted with 'real' triggered landslide event inventories, so we do not use this in the final LRIM model (**Chapter 7**). Nonetheless, the results presented in **Section 6.4** give some interesting insights into spatial clustering and its link to landslide susceptibility, which may be of wider interest to the geomorphic and landslide research communities.

In the following Sections of this **Chapter 6**, we:

6.2 *Describe the two methods used here* to measure spatial clustering.

6.3 *Briefly summarise eight triggered landslide event inventories* analysed in this **Chapter 6**.

6.4 *Present results* from the two methods of measuring spatial clustering described in **Section 6.2** applied to inventories described in **Section 6.3**.

6.5 *Discuss findings and implications* for the Landslide Road Impact Model (LRIM).

6.6 *Summarise* and Conclude.

6.2 Methods to Measure Spatial Clustering

6.2.1 Pair-Correlation

Several methods exist to quantify the spatial clustering of points and polygons in the statistics, spatial analysis and geomorphology literature. Examples include:

(i) *Nearest neighbour functions* which measure the average distance from one point to its nearest point (Cover and Hart, 1967). This technique has been used by Das (2011) in an inventory of landslides in Northern India (described in **Section 6.1** above). They showed that the average distance between landslides was less than what would be expected for a randomly distributed set of points, and thus exhibited spatial clustering.

(ii) *Ripley's K Function* which compares the observed to expected number of points if they were distributed randomly within various radii from a given point. An example of this technique applied to landslides is Lu *et al.* (2014) who compared the count of synthetic aperture radar persistent scatter interferometry points (PSI) within radii of individual landslide centroids, using an inventory of 27,503 landslides occurring across an 8,200 km^2 basin in central Italy, where landslides were separated by type. Lu *et al.* (2014) found that different patterns of spatial

clustering of PSI for different types of landslide, aiding in the differentiation of different landslide types from synthetic aperture radar remote sensing.

(iii) The *Getis-Ord General G* and the *Global Moran's I* statistics which measure whether points with the same magnitude of an attribute value cluster together (Anselin, 1995; Getis and Ord, 1996). This technique was used by Lu *et al.* (2012) as above for the measurement of whether PSI points that were moving at slow or fast speeds cluster 'like with like'. Lu *et al.* (2012) found that PSI points do cluster by movement speed, and that clusters of PSI points moving rapidly indicate movement towards the satellite remote sensor, and that this can be used to identify landslide activity.

(iv) *Kernel density estimation* which counts the number of observations within grid cells of a given size, to give an estimation of density across a region. An example of a geomorphological application of this technique is Maclachlan and Eyles (2013) who measured the kernel density of 3000 drumlins across 1,700 km^2 region of Ontario and linked spatial density of drumlins with different elongation ratios to spatial differences in ice flow direction.

Typically, the above methods work by comparing the pattern of observed points to a randomly generated pattern of points (which will be homogeneously distributed), and measure the amount of deviation between observed and expected values if observed points were distributed randomly. Here, we use a *pair-correlation technique* (Leeuwen and Groeneveld, 1959), selected for its simplicity and ease of implementation. Pair-correlation is a technique to measure the distances between all pairs of points in a 'set', and investigate the probability density distribution of these distances. Pair-correlation has been used across a wide range of disciplines to measure spatial clustering (e.g., Stoyan and Stoyan, 2007 for identifying clusters of tree species; Greenberg *et al.* 2008 for measuring changes in brain neuron activity, Stillinger and Weber, 1984 for measuring the chemical structure of liquids in different thermodynamic states). Examples related to measuring clustering in natural hazards, include:

- *The global distribution of earthquake epicentres*. Using four local and two world-wide catalogues of earthquake locations and magnitudes spanning the years 1969-1975, Kagan and Knopoff (1980) found that the probability density distribution of distances (r) between pairs of earthquake epicentres followed an inverse power-law.

This probability density distribution appeared to be 'generally applicable', regardless of earthquake magnitude or location.

- *The regional distribution of sinkhole occurrence*. Using an inventory of 267 sinkholes occurring across along a highway corridor in Maryland, USA, Zhou *et al.* (2003) used pair-correlation to measure the distances between all pairs of sinkholes. They found that the peak distance between sinkholes is 30 m. This 30 m radius around existing sink holes was proposed to be a useful indicator of where new sinkholes were most likely to form.
- *The planetary distribution of asteroid impact craters*. Turcotte *et al.* (1999) used pair-correlation to measure the spatial clustering of an inventory of 923 craters across the surface of Venus. They showed that the craters did not exhibit clustering (i.e., they were randomly distributed across the surface of the planet). They used this to argue that the surface of Venus had undergone fairly uniform resurfacing in the recent past, as if some parts of the planet's surface were older, there would likely be clusters of more impact craters in these older zones.

The steps used here to calculate the pair-correlation distribution of landslides are as follows:

(i) *Calculate centroids of individual landslide polygons*, based on a weighted average of the polygon's (*N, E*) coordinates.

(ii) *Define study region extent* (A_{Circ}). We define the study region extent as the smallest circle that encompasses all landslide centroids.

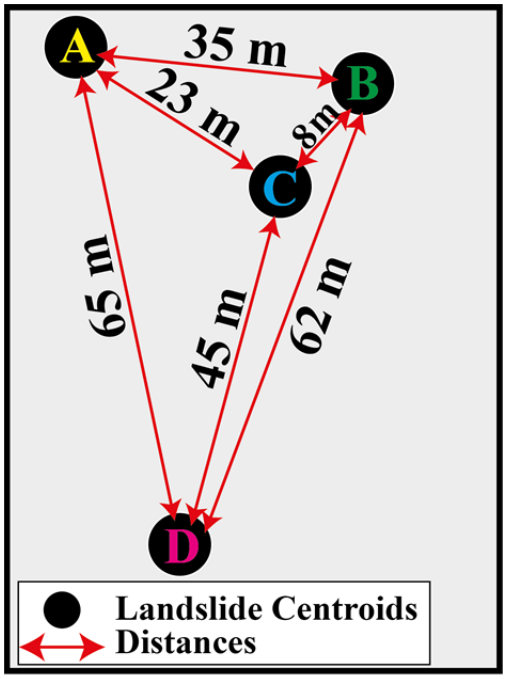
(iii) *Measure linear distance* (*r*) from each landslide centroid to all other landslide centroids, with no 'pair' being repeated; this is illustrated in **Figure 6.2**. The total number of possible pairs measured, is $(N_{LS}-1)!$, where N_{LS} are the number of landslides in the inventory and ! indicates factorial.

(iv) *Analyse probability density distribution* of linear distances *r* between each pair of points (as explained in **Chapter 4** for landslide area and implemented in **Chapter 5** for landslide shape).

(v) *Generate a set of n random Northing (N) and n random Easting (E) coordinates* selected from uniform distributions bounded by the maximum extent of the study region (A_{Circ}), where n = the observed number of landslide centroids in Step (i) (N_{LS}). Repeat steps (iii-iv) for the randomly generated points.

(vi) *Compare the pair-correlation probability density distributions* of the triggered landslide event inventory and the pair-correlation probability density of a randomly distributed set of points in space.

(vii) *[Optional] Normalise by the square root of the study region area ($A_{Circ}^{0.5}$).* Because triggered landslide event inventories span a range of 'magnitudes' in terms of the number of landslides triggered and the extent of the region affected, it may be useful to introduce a linear normalisation in order to compare results from different 'magnitude' inventories. This is done by dividing all distances by $A_{Circ}^{0.5}$.



From	To	Distance (m)
A	B	35
A	C	23
A	D	65
B	C	8
B	D	62
C	D	45

Figure 6.2 *Demonstration of the pair-correlation technique. Left: Four landslide centroids are mapped and the distances r between each the six pairs of points are calculated (r = 8, 23, 35, 45, 62, 65 m). Right: The distances (r) between all points.*

We will show in **Section 6.4.1** that the probability density distribution of a randomly distributed set of point distances (Step v) will follow a distribution that is symmetrical about the mean distance between points. If the triggered landslide event inventory exhibits spatial clustering, the probability density distribution of distances between points will deviate from this this distribution, and may exhibit a skew in the peak of the distribution towards a characteristic distance at which landslides tend to be spaced (Das, 2011). The observed probability density distribution can also be used to calculate the relative probability of observing a landslide at a given distance from a given point. We apply this pair-correlation technique to 8 triggered landslide event inventories, which are described in **Section 6.3** and results are presented in **Section 6.4.1**.

6.2.2 The Kappa-measure

To investigate the relationship between landslide susceptibility and spatial clustering, we developed a grid-based method, which we call the Kappa-measure (κ), The Kappa-measure is the normalised count of an object per km^2 for each cell of a grid. Here, we use the kappa-measure to count the normalised number of landslide centroids per km^2 (κ_{LS}), and in **Chapter 7**, we use the Kappa-measure to count other variables such as the normalised number of road blockages per km^2 (κ_{BL}). In this method, we create a regular grid of 1 km × 1 km cells (each cell has area A_C = 1 km^2) (**Figure 6.3A**) and overlay this with three maps:

- **Map 1.** *Study Area Extent* (A_S), based on the extent of other datasets (Maps 2 and 3). This is different to the study area based on the minimum-bounding circle (A_{Circ}) described in the previous **Section 6.2.1**.
- **Map 2.** *Landslide Susceptibility* (S_{LS}) rasterized to 1 m × 1 m resolution.
- **Map 3.** *Triggered Landslide Event Inventory* (total number of landslide N_{LS}) of triggered landslide centroids.

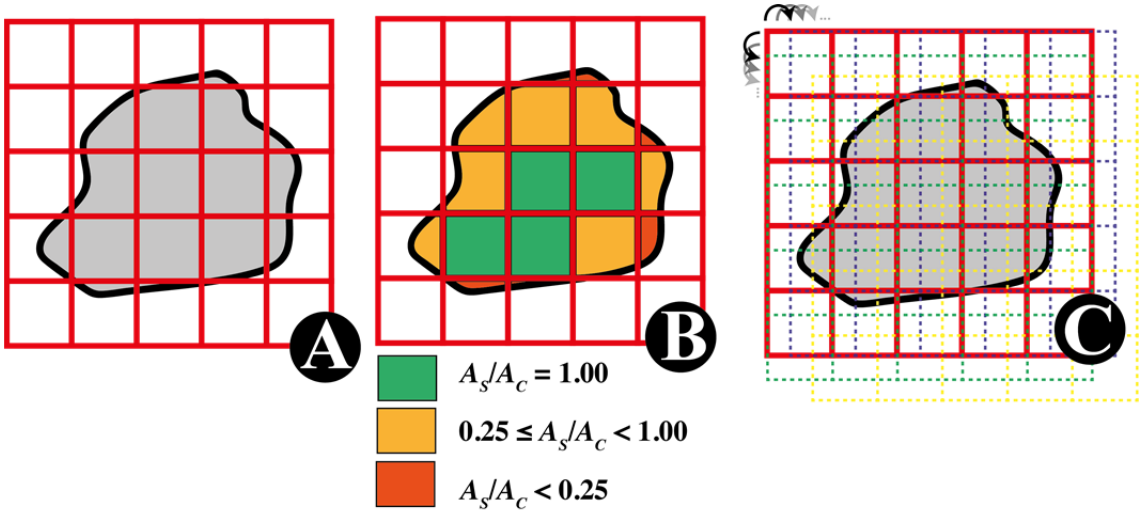


Figure 6.3 *Demonstration of the moving grid. (**A**) A grid of 1 km × 1 km cells (red grid) is overlaid with a study area (grey polygon). (**B**) This grid is moved 200 m in both the x and y directions iteratively in a moving window process. (**C**) The proportion of the cell area (A_C) covered by study region area (A_S) is calculated (A_S/A_C). If $A_S/A_C < 0.25$, the results for this cell are rejected; if $0.25 \leq A_S/A_C < 1.00$, the results for this cell are normalised to the total area of the cell. If $A_S/A_C = 1.00$, the results for this cell are recorded with no modification.*

For each 1 km × 1 km grid cell which we give index i = 1, 2, 3,…, N_{cells}, where N_{cells} are the total number of cells required to cover the study area extent, we calculate the following:

(i) *Proportion of each cell's area (A_C) covered by study region area* (A_S) is calculated for each cell i, giving (A_S/A_C). This accounts for irregularly shaped study regions where the edge cells may not be entirely covered by study area. If (A_S/A_C) < 0.25, i.e.

<25% of the cell is covered by study area, the cell is rejected from further analysis (**Figure 6.3B**).

(ii) *Landslide susceptibility (S_{LS}) summary statistics for all cells,* (e.g., minimum, median, maximum, interquartile range of S_{LS}) based on raster landslide susceptibility map cell values.

(iii) *Number of landslide centroids in a given cell (n_{LS}).*

(iv) *Number of landslide centroids normalized to be a 'full' 1 km × 1 km cell:* $\kappa_{LS[i]} = n_{LS[i]}(A_{C[i]}/A_{S[i]})$. In other words we have normalised the number of landslide centroids in cell *i* ($n_{LS[i]}$) (Step iii) by the proportion of the cell *i* that is covered by the study area ($A_{S[i]}/A_{C[i]}$) (Step i). We will call this final variable κ_{LS} the Kappa-measure for each cell.

We perform the above method for each cell in the grid of 1 km × 1 km cells, but want the entire grid to move as a fraction of a cell, thus allowing for the 1 km × 1 km cells to overlay different parts of the susceptibility and landslide inventory maps. Therefore, we initially start the grid such that the lower left 1 km × 1 km cell's outside corner lower corner is at (N = 0 m, E = 0 m), but then we shift the entire grid of 1 km × 1 km cells 200 m both horizontally and vertically, in all possible configurations, as a moving window process, resulting in 25 different shifts of the grid of cells (**Figure 6.3C**). For each of the N *cells* in the grid, and the 25 total different overlays of the grid of cells, we repeat Steps (i) to (iv), calculating for every cell *i* the Kappa-measure κ_{LS}. Overall, for a given triggered event inventory, we calculate a total number of Kappa-measures of $25 \times N_{\text{cells.}}$.

We calculate this Kappa-measure (for each 1 km × 1 km cell, and each shifting of the grid by 200 m) for two of the triggered landslide event inventories (Collazzone and Northridge: Oat Mountain) described in the following **Section 6.3** and present results of the Kappa-measure in **Section 6.4.2**. The locations of the landslides in these two inventories (Collazzone, Italy and Northridge: Oat Mountain, California, USA) are also sites where the Landslide Road Impact Model (LRIM) is developed and confronted against real data, which will be shown in **Chapter 7**. In **Chapter 7**, we also show the use of the Kappa-measure to explore the relationship between road presence and landslide occurrence on a cell-by-cell basis.

6.3 Triggered Landslide Event Inventories Analysed

In **Chapters 2, 4** and **5** we have discussed the importance of using substantially complete triggered landslide event inventories to avoid bias towards particular types and sizes of landslide when performing statistical analysis. This principle also applies to the analysis of spatial clustering performed here. For the pair-correlation technique, we use eight triggered landslide event inventories of varying 'confidence' in terms of their completeness. These inventories are described in **Table 6.1**. We rank the inventories used here as low, medium or high confidence based on Harp *et al.* (2011)'s criteria for a substantially complete triggered landslide event inventory, which was outlined in **Chapter 2**, **Table 2.5**. Note, some of these inventories differ to those used in **Chapter 5**, based on what was available and/or suitable at the time of performing this research. For the Kappa-measure, we analyse two of these eight inventories (Collazzone, Italy and Northridge: Oat Mountain, California, USA), as these are two sites where landslide susceptibility maps are available.

Table 6.1 *Triggered landslide event inventories used to measure patterns of spatial clustering in* ***Chapter 6****. Inventories are rated in terms of 'confidence' (low, medium, high) in how substantially complete they are, based on Harp et al. (2011)'s criteria, introduced in* ***Chapter 2****. Study region size corresponds to the area of the smallest circle encompassing all landslide centroids.*

Inventory Name (reference)	Cluster Analysis Techniques Applied Here	# Land-slides (N_{LS})	Trigger	Study region size (km²) (A_{Circ} or A_S)	Notes	Inventory Confidence
Collazzone (Cardinali *et al.*, 2000)	(a) Pair-Correlation (b) Kappa-measure	413	January 1997 Snowmelt in Collazzone, Italy	(a) A_{Circ}= 201 (b) A_S =79	**Broadly low mobility** **Subset of Umbria, Italy dataset (also analysed here)** 62% small shallow slips Some larger earth flows and deep seated slides Occurring on slopes > 20° with shear surface parallel to slope	**Medium confidence:** Low number of landslides. Considered significantly 'complete' (e.g., Malamud *et al.,* 2004a).

Inventory Name (reference)	**Cluster Analysis Techniques Applied Here**	**# Land-slides (N_{LS})**	**Trigger**	**Study region size (km²) (A_{Circ} or A_S)**	**Notes**	**Inventory Confidence**
Guatemala Mitch (Bucknam *et al.,* 2001)	Pair-Correlation	9,594	1998 Hurricane Mitch in Guatemala	A_{Circ}= 35,968	**Significant proportion high mobility landslides** Many small shallow translational slides, many mobilised into debris flows Some large translational slides which transitioned into debris flows following channels Occurring on moderate to steep hillslopes across diverse geology.	**High confidence:** High number of landslides. Considered to be 95 % complete (Bucknam *et al.* 2001).
Guatemala Earthquake (Harp *et al.,* 1981)	Pair-Correlation	6,402	1974 Earth-quake, Guatemala City, Guatemala	A_{Circ} = 11,618	**Significant proportion high mobility landslides** Generally rockfalls and debris slides. 90% of landslides located in Pleistocene pumice deposits. Landslides location controlled by pre-existing fractures, seismic intensity and amplification and slope angle.	**Low confidence:** High number of landslides. Created from 1:50,000 imagery covering majority of affected region. Some landslides mapped that occurred before earthquake.

Inventory Name (reference)	**Cluster Analysis Techniques Applied Here**	**# Land-slides (N_{LS})**	**Trigger**	**Study region size (km²) (A_{Circ} or A_S)**	**Notes**	**Inventory Confidence**
Liguria (Mondini, 2014)	Pair-Correlation	537	2011 heavy rainfall in Liguria, northwest Italy	A_{Circ} = 34	**Generally high mobility landslides** All shallow landslides Translational slides, earth flows, soil slips and rotational slides. Slides located at the foot of slopes Flows channelled in secondary streams.	**Low confidence:** Medium number of landslides. Created from semi-automatic interpretation of satellite imagery. Some small landslides missed (~8.7% compared to photo interpretation).
Messina (Mondini *et al.*, 2011b)	Pair-Correlation	1,670	2009 rainfall in Messina, Southern Italy	A_{Circ} = 24	**Generally high mobility landslides** Mainly shallow soil slides and debris flows, many reactivations. Occurred in clusters on open slopes and drainage channels. Many caused by poor drainage of road network.	**Medium Confidence:** Medium number of landslides. Produced from semi-automatic detection in high resolution remote sensing imagery. 86% match to ortho-photograph inventory.
Northridge (Harp and Jibson 1995)	(a) Pair-Correlation (b) Kappa-measure	(a) 11,111 (b) 1,356	1994 Northridge earthquake, California, USA	(a) A_{Circ}= 6,648 (b) A_S = 156	**Generally low mobility landslides** Many shallow, highly disrupted falls and slides Occurred in weakly cemented sediments. Most larger slides travelled < 50 m, some travelled up to 200 m.	**High Confidence:** High number of landslides. Considered significantly complete (e.g., Malamud *et al.,* 2004a).

Inventory Name (reference)	**Cluster Analysis Techniques Applied Here**	**# Land-slides (N_{LS})**	**Trigger**	**Study region size (km²) (A_{Circ} or A_S)**	**Notes**	**Inventory Confidence**
Taiwan Morakot (Mondini *et al.,* 2011a)	Pair-Correlation	1,466	2008 Typhoon Morakot, southern Taiwan	A_{Circ} = 115	**Significant Proportion High Mobility Landslides** Mainly shallow landslides Some debris flows	**Medium Confidence:** Medium number of landslides. Produced from semi-automatic detection from 2m × 2m remote sensing imagery. 91.3% agreement with ortho-photography inventory.
Umbria (Cardinali *et al.,* 2000)	Pair-Correlation	4,222	1997 snowmelt, Umbria, central Italy	A_{Circ} = 9,641	**Broadly low mobility** 62% small shallow slips Some larger earth flows and deep seated slides Occurring on slopes > 20° with shear surface parallel to slope	**High Confidence:** High number of landslides. Considered significantly 'complete' (e.g., Malamud *et al.,* 2004a).

The inventories in **Table 6.1** represent a variety of:

1. *Global locations and topographic settings*: 4 inventories from Italy, 2 inventories from Guatemala, 1 inventory from the USA and 1 inventory from Taiwan.
2. *Triggering mechanisms*: 4 rainfall-triggered inventories, 2 snowmelt-triggered inventories (one is a subset of the other) and 2 earthquake-triggered inventories.
3. *Typical landslide type*: 5 inventories are dominated by 'high mobility' type landslides (e.g., many of the landslides are debris flows and rock falls) and 3 inventories are dominated by 'low mobility' landslides (e.g., many of the landslides are rotational or translational slides).
4. *Confidence in the robustness of the inventory*: 3 inventories are considered 'high confidence' due to large numbers of landslides and well-documented

methodologies explaining how they are substantially complete, 3 inventories are considered 'medium confidence' due to either smaller numbers of landslides or less confidence in the completeness of the inventory and 2 inventories are considered 'low confidence' due to known issues with completeness.

We hope that by investigating a variety of settings (outlined in the four points above) that we can test whether there is any 'generally applicable' behaviour in spatial clustering of landslides. The 'medium' and 'low' confidence inventories must be treated with more caution for statistical analysis, but also represent the opportunity to test the methods outlined in **Section 6.2** with datasets that reflect what is more typically available for more regions of the world.

In the following **Section 6.4.1** we present the results of the pair-correlation technique applied to all eight triggered landslide event inventories listed in **Table 6.1**. In **Section 6.4.2** we show the results of the Kappa-measure for the Collazzone and Northridge inventories.

6.4 Results

6.4.1 Results: Pair-Correlation Measure of Landslide Spatial Clustering

6.4.1.1 Pair-correlation Results by Landslide Inventory

The pair-correlation methodology described in **Section 6.2.1** was applied to eight triggered landslide event inventories, which vary in data confidence, number of landslides triggered, triggering mechanism and extent of the region affected (described in **Table 6.1**). We compare the spatial clustering of observed landslide centroids to randomly generated points over the same region (which will be homogeneously distributed) to test whether landslide centroids exhibit clustering or are evenly distributed. An example is shown in in **Figure 6.4C**, where we observe that the pair-correlation distribution of randomly generated points is approximately symmetrical, and peaks at r = 3,000 m. This is because landslides are fairly evenly spaced across the circular study region (**Figure 6.4B**). Whereas, the pair correlation distribution for observed landslide centroids is more 'noisy' and skewed towards smaller values of r, peaking at r = 2,000 m. The difference between the observed and random points pair correlation distribution suggests that observed landslides exhibit clustering.

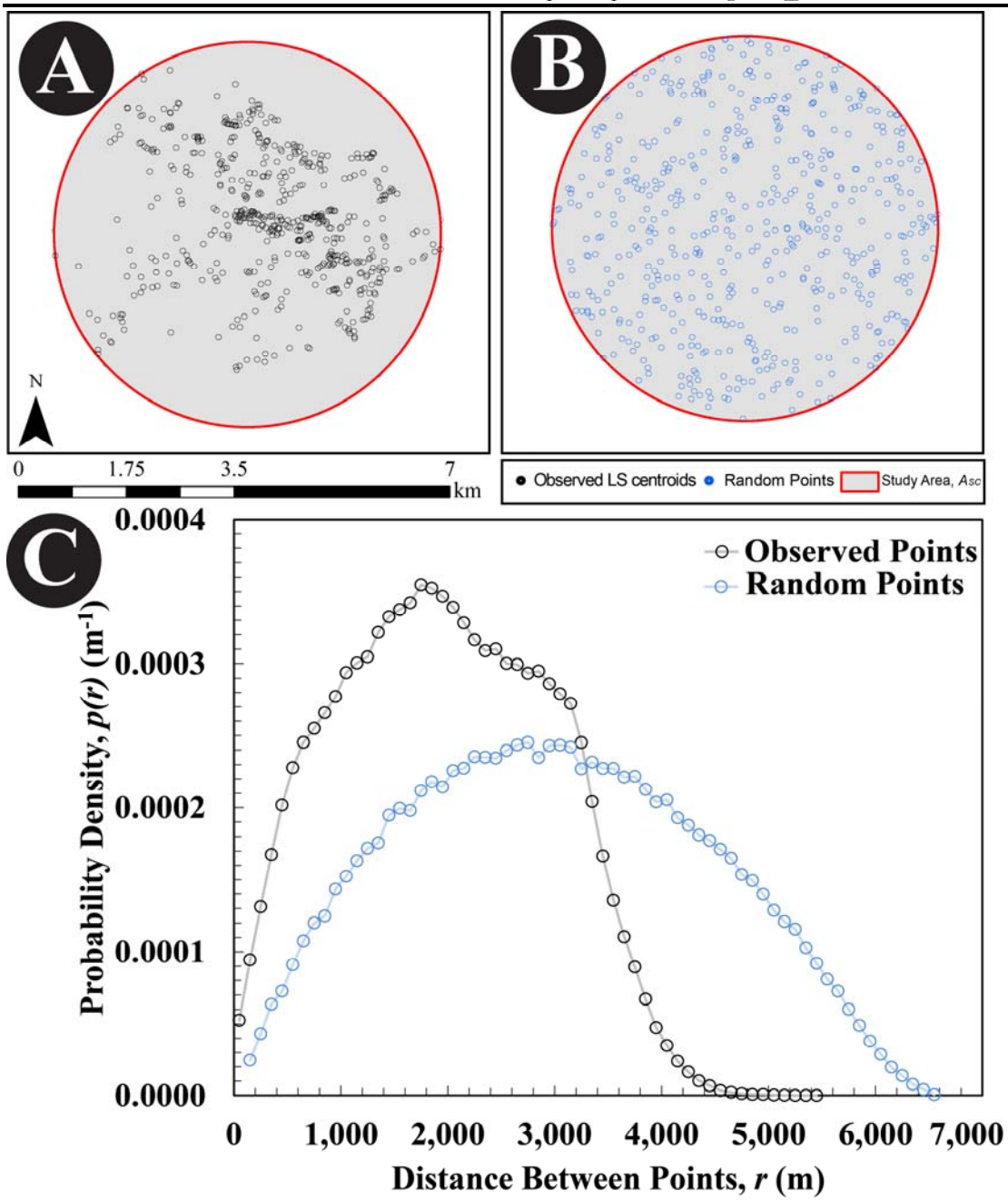


Figure 6.4 *Example of the spatial distribution of* ***(A)*** *observed landslide centroids for the Liguria, Italy inventory over the minimum area circle that encloses all points (A_S) and* ***(B)*** *n randomly generated points over the same study area (A_S), where n = observed number of landslide centroids. In this case, n = 537* ***(C)*** *Comparison of the probability density distribution of pair-correlation distances between points (r) for observed centroids and randomly generated points.*

The probability density distributions for all eight inventories examined (and corresponding sets of *n* randomly generated points) in **Figure 6.5** are shown on logarithmic *x* and *y* axes due to the large span in distances between landslides.

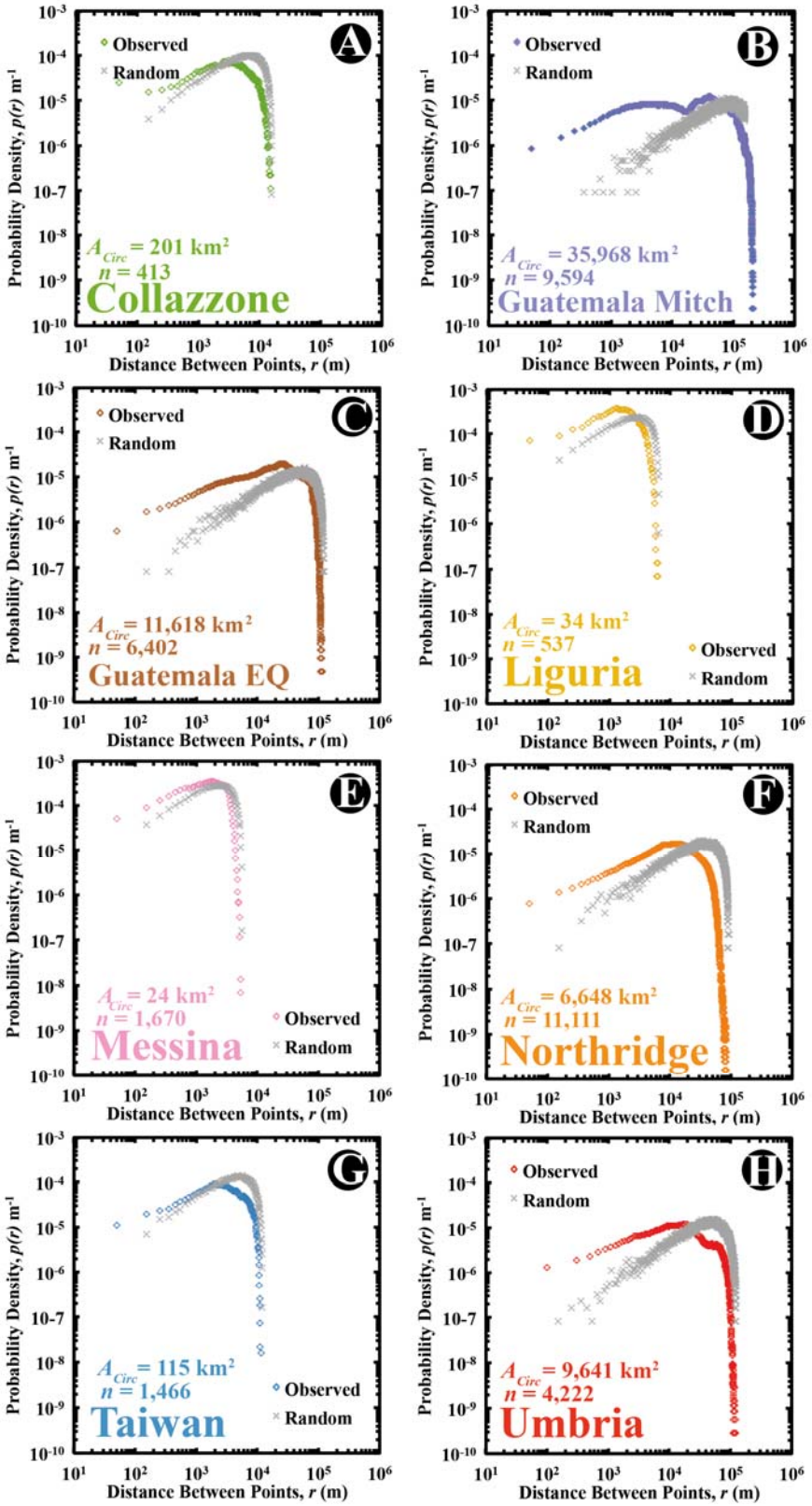


Figure 6.5 *Probability density distribution of distances between all landslide centroids (r) in eight different triggered landslide event inventories (**A**) Collazzone, Italy (Cardinali, 2000), (**B**) Guatemala Hurricane Mitch (Bucknam et al., 2001), (**C**) Guatemala City earthquake (Harp et al., 1981), (**D**) Liguria, Italy (Mondini, 2014), (**E**) Messina, Italy (Mondini, et al., 2011b)(**F**) Northridge, USA (Harp and Jibson, 1995), (**G**) Taiwan Typhoon Morakot (Mondini et al., 2011a) (**H**) Umbria, Italy (Cardinali et al., 2000). For each inventory, the observed results are compared to the probability density distribution of points which are randomly generated over the same study region (grey cross markers). The inventories are described in full in* ***Table 6.1****.*

6.4.1.2 Similarities and Differences in Pair-Correlation between Inventories

In **Figure 6.5**, we see the distribution of probability density of the pair-correlation distances *r* differs between observed and random points for all eight inventories examined. For all inventories other than Taiwan (**Figure 6.5G**), the probability density distribution of observed point pair-correlation distances (*r*) is skewed towards smaller distances and the gradient of the left hand tail (small distances between points) is shallower, suggesting that landslides exhibit clustering towards smaller spatial scales. The right hand tails of the distributions appear to follow a similar gradient for both observed and random points, although observed distributions tend to 'die off' at a slightly slower rate, meaning that there are a small number of observations of very large distances between points. For each observed inventory, the shape of the probability density distribution appears to be similar. We show all observed probability density distributions for the eight triggered landslide event inventories on one plot in **Figure 6.6**.

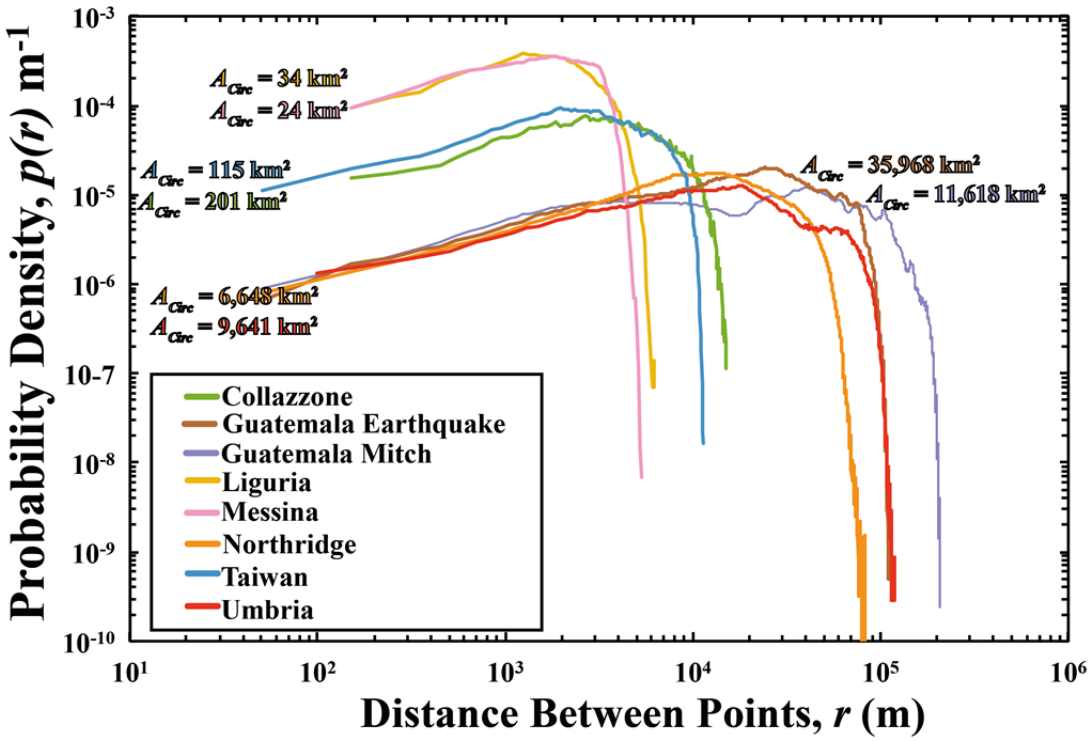


Figure 6.6 *Observed probability density distributions of distances (r) between all landslide centroids in each of eight triggered landslide event inventories examined (described in* ***Table 6.1****). The area of the circular study region (A_S) is also included near each line.*

Figure 6.6 illustrates the similarity in the shape of probability distributions of pair-correlation distances (*r*) between landslide centroids for all triggered landslide event inventories examined. The position of the probability distribution along the *x*-axis is linked to the size of the study region (A_{Circ}): the larger the study region, the greater the probability

that some landslides will be very far apart if they are on the periphery of the study region. To further investigate similarities in the pair-correlation between inventories, we normalised all distances (r) by the square root of the respective study region area ($A_{Circ}^{0.5}$), with results shown in **Figure 6.7**.

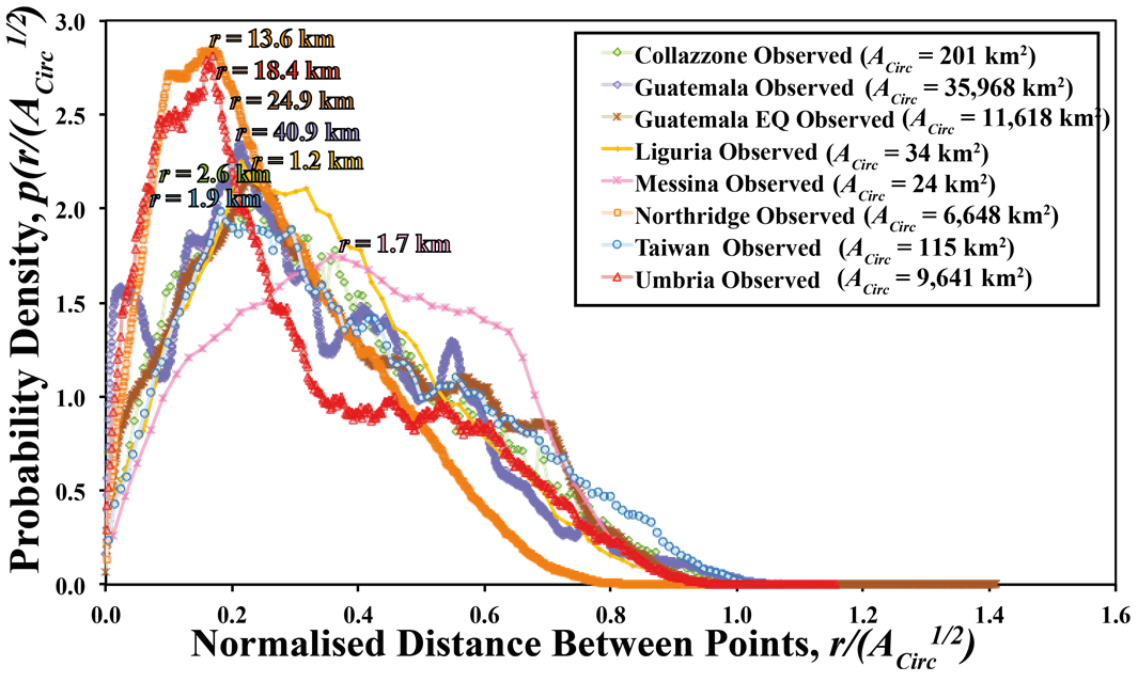


Figure 6.7 *probability density distributions of normalised distances between landslide centroids in each of eight triggered landslide event inventories examined (outlined in* ***Table 6.1****). All distances are normalised by the square root of the study area (A_{Circ}). Values of A_{Circ} are included in the legend. The (non-normalised) distance r corresponding with the maximum probability for each distribution is also given above each curve.*

When pair-correlation distances r between landslide centroids are normalised by the square root of the size of the study region (A_{Circ}), we observe that although the general shape of probability density distributions is approximately similar between inventories, there is some variability regarding where the peak of the distributions occur and the 'smoothness' of the distributions. There appears to be a distinction in where the peak of the distribution occurs between:

(i) *Low mobility triggered landslide event inventories* (Northridge Earthquake and Umbria Snowmelt) where normalised distances between points ($r/A_{Circ}^{0.5}$) tend towards smaller values and thus landslides are proportionally more clustered. These inventories also happen to be 'high confidence' in terms of completeness and large numbers of observations.

(ii) *High mobility, triggered landslide event inventories* (Guatemala Hurricane Mitch, Liguria rainfall, Messina rainfall, Taiwan Typhoon Morakot and to a certain extent Guatemala City Earthquake which was dominated by rockfalls and debris flows). In

these inventories, the normalised distances between points ($r/A_{Circ}^{0.5}$) tend towards slightly larger values, and peaks at larger values, suggesting proportionally greater spacing (less clustering) between landslides. This could potentially be linked to dominance of debris-flow type landslides which tend to follow channel morphology (Jakob and Hungr, 2005) which may exhibit a different spatial pattern of clustering.

The 'smoothness' of the probability density distributions also appears to vary between inventories: inventories with large numbers of observations and occurring over large study areas (A_{Circ}) (Northridge Earthquake, Umbria Snowmelt, Guatemala Earthquake and Guatemala Hurricane Mitch) tend to be 'noisier' and exhibit peaks and troughs. This could potentially be attributed to the structure of the landscape across a study region, in terms of topography, geology, land use and other factors controlling landslide susceptibility. For example, across a larger area there may be a greater the probability that this landscape includes areas where no landslides occur (e.g., lakes, flat areas, urban areas with slope protection measures). In **Section 6.4.2** we investigate this in more detail by looking at the Kappa-measure of number of landslides per square kilometre in relation to landslide susceptibility.

6.4.1.3 Comparison of Observed and Landslide-Road Impact Model (LRIM) modelled Pair-Correlation Distribution Inventories

We now briefly compare the pair-correlation of observed inventories to synthetic triggered landslide event inventory maps created in the Landslide-Road Impact Model (LRIM) (which will be described fully in **Chapter 7**, but the process was summarised in **Section 6.1**). This analysis is applied to a subset of the 1,356 landslides in Northridge triggered landslide event inventory (the Oat Mountain Quadrangle) where the other LRIM model inputs are available (landslide susceptibility map, study region extent, road network map, digital elevation model and broad characteristics of road-landslide protection). In **Figure 6.8** we show a comparison of the probability density distribution of pair-correlation distances (r) for (i) 1,356 observed landslide centroids for Northridge triggered landslide event inventory, (ii) 1,356 LRIM modelled landslide centroids (random points, but then conditioned on underlying susceptibility available for Northridge region, described in more detail in Section 6.1 and **Chapter 7**) and (iii) 1,356 randomly distributed points (without conditioning based on underlying susceptibility or any other LRIM model inputs). All three are done for the same study extent area.

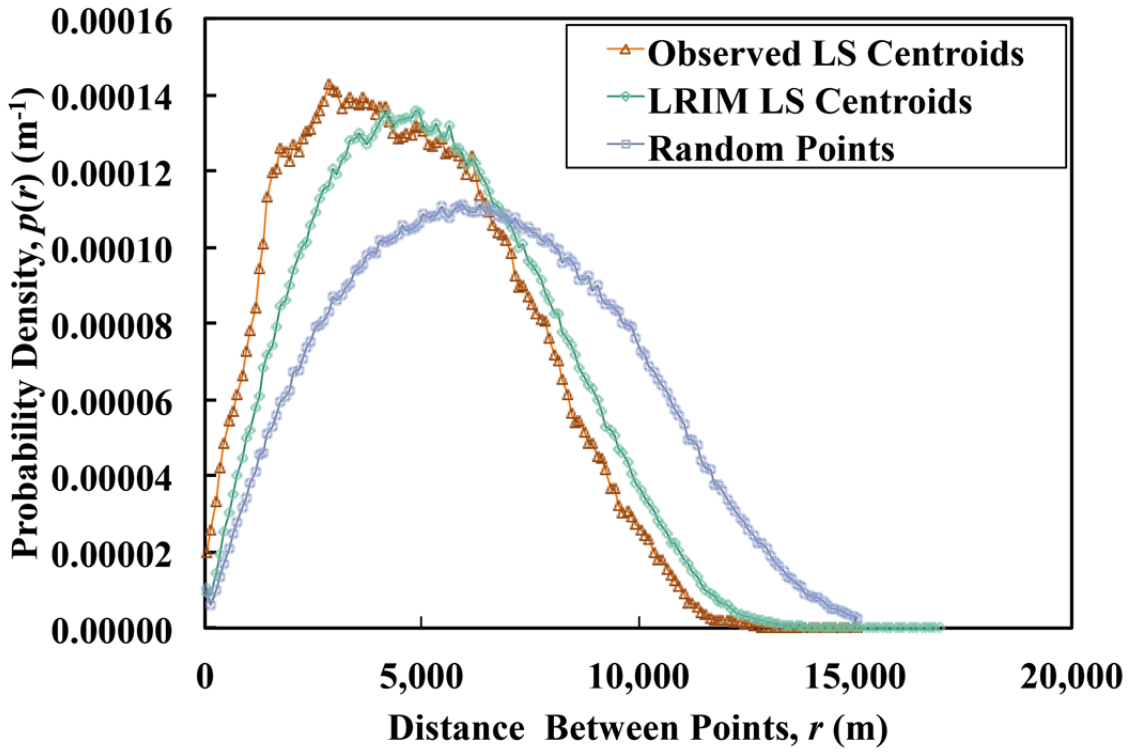


Figure 6.8 *Comparison of probability density distribution of distances between landslide centroids (i) observed in the inventory of 1,356 landslides triggered by the Northridge Earthquake in the Oat Mountain Quadrangle (subset of Northridge inventory) (orange markers), (ii) from the synthetic triggered landslide event inventory map created in the Landslide-Road Impact Model (LRIM) for the Oat Mountain Quadrangle (green markers) and (iii) randomly distributed points in space (blue markers).*

Figure 6.8 illustrates the disparity in the clustering of the 1,356 landslide centroids observed, those created in the LRIM, and those randomly distributed. Both the observed and modelled (LRIM) landslide inventory maps exhibit clustering, as they do not follow the distribution of randomly generated points. However, the observed points exhibit a finer scale of clustering than those created by LRIM (i.e., landslides tend to occur in closer proximity to one another). This suggests that the landslide susceptibility map used to condition where the landslide centroids occur in LRIM (described in **Section 6.1** and fully in **Chapter 7**) does not recreate the spatial pattern of clustering found in the actual inventory. In the following **Section 6.4.2** we further investigate the relationship between landslide susceptibility and spatial clustering of landslides.

6.4.2 Results: Kappa-measure of Landslide Spatial Clustering

6.4.2.1 Number of Landslide Centroids per Square Kilometre

In this **Section 6.4.2**, we present the results of analysing the normalised number of landslide centroids per square kilometre grid cell (the Kappa-measure, κ_{LS} in a moving-window type approach, methodology described in **Section 6.2.2**). The Kappa-measure was

applied to the Collazzone inventory of 413 landslides triggered by snowmelt in 1997 (this is a subset of the Umbria snowmelt inventory, described in **Table 6.1**) and the Oat Mountain inventory of 1,356 landslides triggered by the 1994 Northridge Earthquake (this is a subset of the Northridge inventory, described in **Table 6.1**). These regions were selected for analysis as landslide susceptibility maps are available for these sites, allowing investigation of the link between spatial clustering and landslide susceptibility (shown later in **Section 6.4.2.2**). In **Figure 6.9**, we show the probability density distribution of κ_{LS} for Collazzone and Oat Mountain. To aid visualisation, **Figure 6.9A** and **B** are presented on logarithmic axes. Consequently, cells where there are no landslide centroids ($\kappa_{LS} = 0$) are plotted as $\kappa_{LS} = 10^{-1}$.

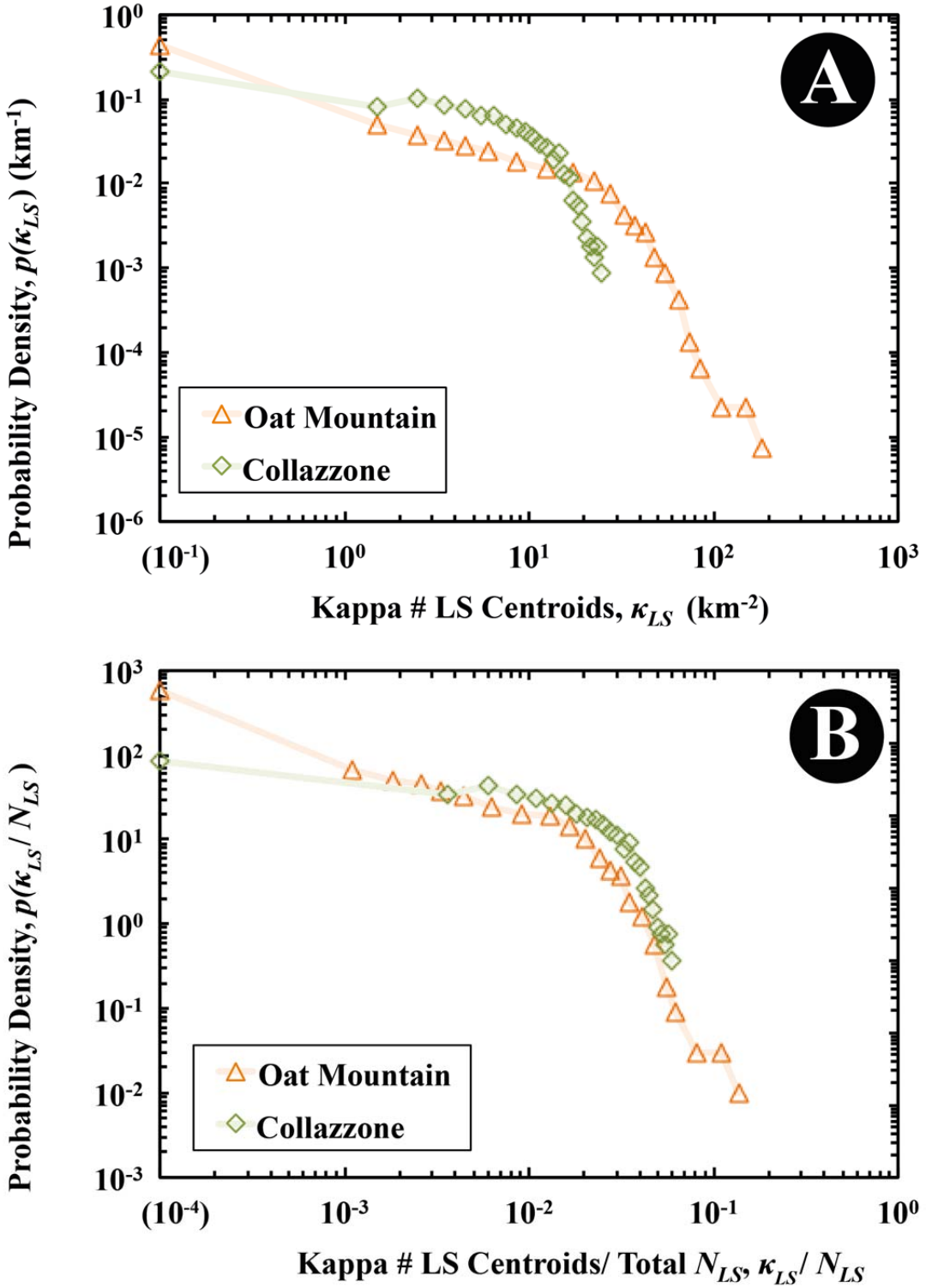


Figure 6.9 *Kappa-measure (κ_{LS}) of normalised number of landslide centroids km^{-2} for Collazzone and Oat Mountain (subset of Northridge inventory).* ***(A)*** *Standard Kappa-measure* ***(B)*** *Kappa-measure is normalised by the total number of landslides in the inventory, to allow direct comparison between locations. The Collazzone inventory has 413 landslide centroids over a 79 km^2 study region. A grid with a total of 181 (1 km × 1 km) cells is required to cover the entire study region. When using the moving window approach (cells moved 200 m at a time in both N and E directions, see* ***Section 6.2.2*** *for discussion) this equates to 2,231 (1 km × 1 km) cells. The Oat Mountain inventory includes 1,356 landslides (subset of 11,111 landslides triggered by the Northridge Earthquake) over a 156 km^2 study region. A grid with a total of 182 (1 km × 1 km) cells is required to cover the entire study region. When using the moving window approach, this equates to 4,517 (1 km × 1 km) cells.*

Figure 6.9A shows that the probability density of κ_{LS} is highest at $\kappa_{LS} = 0.1$ and gently decays between $0.1 < \kappa_{LS} < 10$. After this point, κ_{LS} dies off more rapidly as a near-straight line on logarithmic axes up to values of $\kappa_{LS} \approx 20$ for Collazzone and $\kappa_{LS} \approx 110$ for Oat Mountain. In practical terms, this means that there is a high probability ($p(\kappa_{LS}) = 0.2$–0.4) of observing a cell containing no landslide centroids, there is a medium probability ($0.1 < p(\kappa_{LS}) < 0.01$) of observing cells containing a small number of landslides ($1 < \kappa_{LS} < 10$ ls km^{-2}) and then a relatively low probability of cells containing a large number of landslide centroids (where $\kappa_{LS} > 10$ ls km^{-2}). The near straight line of the distribution for values of $\kappa_{LS} > 10$ suggests some potential inverse power-law behaviour for large values of κ_{LS}. The fact that κ_{LS} spans four orders of magnitude ($0.1 < \kappa_{LS} < 100$) suggests a high degree of clustering: many cells contain no landslide centroids and a smaller number of cells contain large numbers of landslides.

The Collazzone and Oat Mountain inventories vary in 'magnitudes' of each inventory in terms of total number of landslides (N_{LS}), and average landslide density across the entire study area (L_D). To more directly compare the probability density distributions of κ_{LS} between the locations, in **Figure 6.9B** we normalise κ_{LS} per cell by the total number of landslides in the inventory (κ_{LS} / N_{LS}). κ_{LS}/N_{LS} is effectively a measure of the proportion of the total landslide inventory contained in each cell. In **Figure 6.9B**, we observe relatively similar behaviour in the probability density distribution $p(\kappa_{LS}/N_{LS})$ for both Collazzone and Oat Mountain, with a gentle decay in $p(\kappa_{LS}/N_{LS})$ for $10^{-4} < \kappa_{LS}/N_{LS} < 3 \times 10^{-2}$ (i.e., < 3% of the total number of landslides in the inventory occur within an individual 1 km^2 cell). After this point, the decay is more rapid, for $3 \times 10^{-2} < \kappa_{LS}/N_{LS} < 10^{-1}$ (i.e., 3–10 % of the total landslide inventory occurs within an individual 1 km^2 cell). There does appear to be a difference in the probability density of cells where $\kappa_{LS}/N_{LS} = 10^{-4}$. This means that the probability of observing a cell with no landslide centroids is approximately five times higher for Oat Mountain than Collazzone. This may be attributable to a significant area of urban development on relatively low susceptibility colluvium in the South West of the Oat Mountain study area, whereas the Collazzone study area is comparatively undisturbed (a small number of hill towns are evenly distributed across the area), shown in **Figure 6.10**. This adds weight to the argument in **Section 6.4.1** that landscapes may contain low landslide susceptibility 'patches', effectively disrupting the pattern of spatial correlation over a wide region. We now briefly investigate the link between κ_{LS} and landslide susceptibility for Collazzone (Italy) and Oat Mountain (USA).

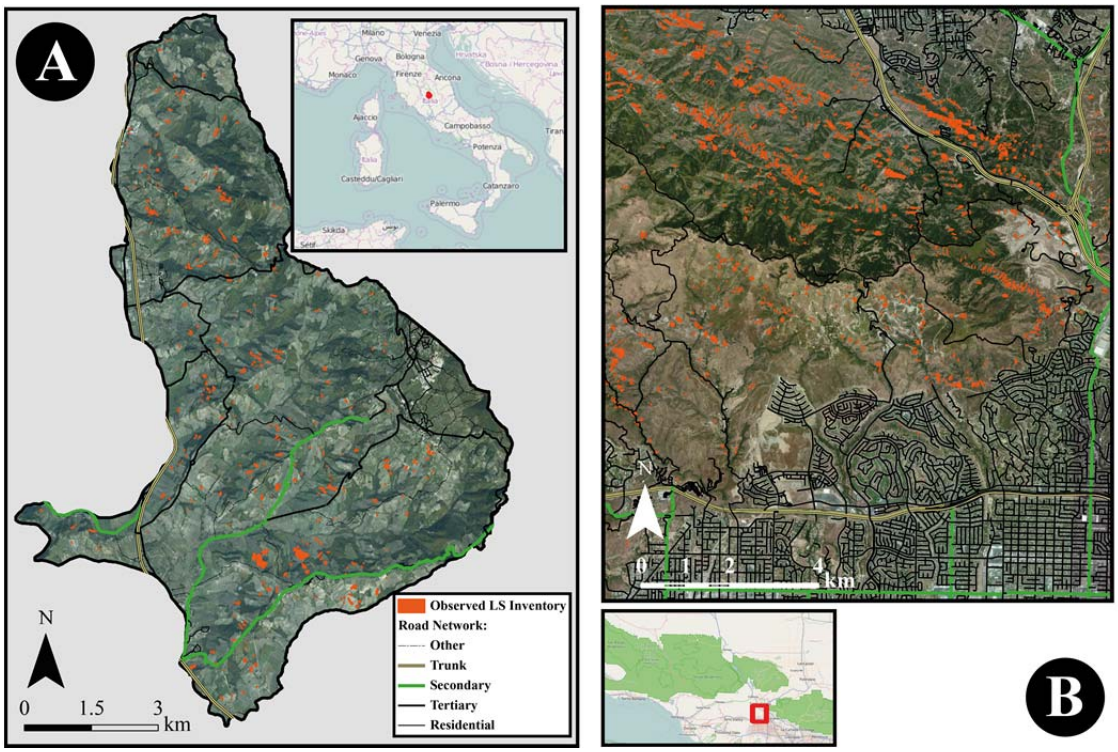


Figure 6.10 *Maps of observed triggered landslide event inventories for* ***(A)*** *Collazzone (Italy) and* ***(B)*** *Oat Mountain (USA) overlaid with aerial imagery and road network maps to illustrate the potential link between landslide spatial clustering and landscape structure. Base imagery: ESRI (2015), Collazzone road map: OSM (2014), Oat Mountain road map: USCB (2006).*

6.4.2.2 Link Between Kappa-measure (κ_{LS}) and Landslide Susceptibility (S_{LS})

In this **Section 6.4.2.2**, we investigate the link between spatial clustering of landslides, measured using the normalised number of landslide centroids per square kilometre (κ_{LS}) and landslide susceptibility (S_{LS}) for the Collazzone (Italy) and Oat Mountain (USA) study regions. Landslide susceptibility is defined as the probability of observing a landslide within a zone over the time period in which the geo-environmental conditions remain the constant within that zone (e.g., Guzzetti *et al.* 1999; 2006a). The two susceptibility maps used here are introduced more fully in the following **Chapter 7**, but we briefly outline them here:

- *The Collazzone (Italy) shallow landslide susceptibility map* was produced by Rossi *et al.* (2010a) at the slope unit scale through linear discriminant analysis of 51 variables including terrain and land cover, and compared to a multi temporal inventory of 2,455 landslides that occurred between 1941 and 1996 as a result of various triggers. They then validated this against a 9-year multi temporal inventory of landslides from 1997–2005.

- *The Oat Mountain (California, USA)* 10 m × 10 m spatial resolution raster susceptibility map was produced by Jibson *et al.* (1998). It was based on an infinite slope model and inventory of landslides triggered by the Northridge Earthquake. Susceptibility is split into 7 discrete classes. This map is available as a non-georeferenced tiff file (Jibson *et al.*, 1998), which we georeferenced manually from visual identification of ground control points – there were some small errors in spatial alignment.

We acknowledge that the concept, calculation and presentation of landslide susceptibility maps for Collazzone and Oat Mountain differ, so results may not necessarily be directly comparable.

Figure 6.11 shows the probability density distribution of κ_{LS} classified into categories of median landslide susceptibility within that cell (S_{LS}) for Collazzone (Italy) and Oat Mountain (California, USA). Although the calculation of S_{LS} differs between the two locations, the categories of S_{LS} are approximately equivalent between the two sites.

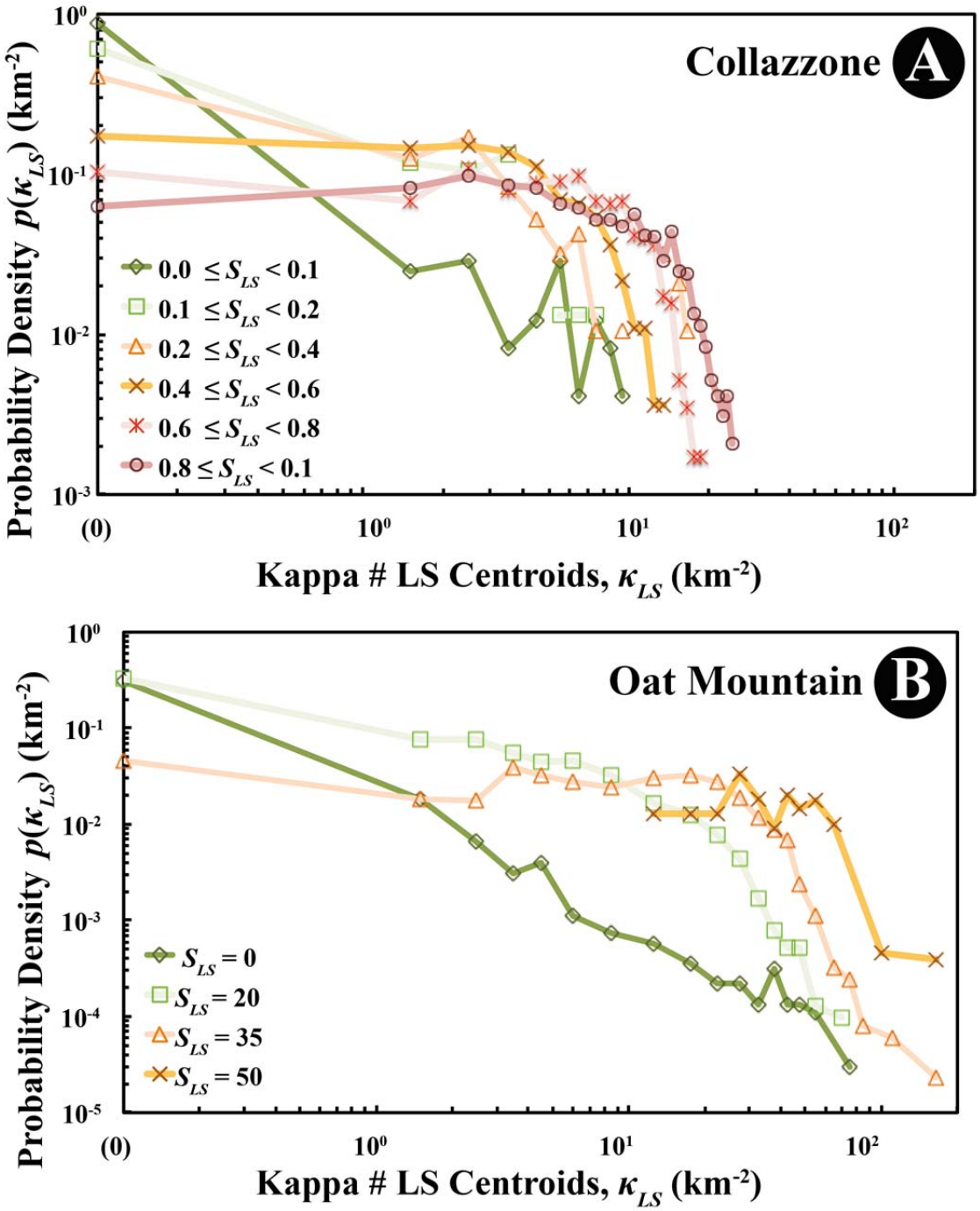


Figure 6.11 *Probability density of the Kappa-measure (κ_{LS}) of normalised number of landslide centroids km^{-2} categorised by median landslide susceptibility in cells (S_{LS}) for* ***(A)*** *Collazzone (Italy) and* ***(B)*** *Oat Mountain (subset of Northridge inventory, California, USA). The Collazzone inventory has 413 landslide centroids over a 79 km^2 study region. A grid with a total of 181 (1 km × 1 km) cells is required to cover the entire study region. When using the moving window approach (cells moved 200 m at a time in both N and E directions, see* ***Section 6.2.2*** *for discussion) this equates to 2,231 (1 km × 1 km) cells. The Oat Mountain inventory includes 1,356 landslides (subset of 11,111 landslides triggered by the Northridge Earthquake) over a 156 km^2 study region. A grid with a total of 182 (1 km × 1 km) cells is required to cover the entire study region. When using the moving window approach, this equates to 4,517 (1 km × 1 km) cells.*

Figure 6.11 shows that for both sites, the probability density distribution shifts towards larger values of κ_{LS} for higher classes of S_{LS} (i.e., a greater proportion of observations have

high values of κ_{LS}) This seems fairly logical there is a higher probability of observing a greater number of landslide cells in a high susceptibility cell. What may seem less intuitive is that there is still the possibility of observing cells with very low values of median landslide susceptibility that contain a cluster of landslides (even though the probability is lower). This is more notable in the Collazzone inventory (**Figure 6.11A**), where the probability distribution for each class of landslide susceptibility extends towards relatively high values of κ_{LS} (i.e., it is possible to observe values of $\kappa_{LS} = 10$ LS km^{-2} for cells where the median susceptibility is $0.0 \leq S_{LS} < 0.1$). Conversely, there is a relatively high probability of observing cells where median S_{LS} is very high, but no landslides are observed.

The Kappa-measure results are calculated at 1 km × 1 km cell resolution, but landslide susceptibility tends to vary over a finer spatial scale than this. To investigate whether the findings shown in **Figure 6.11** are simply the result of high variability in S_{LS} within a 1 km^2 cell, in **Figure 6.12**, we show a plot of median landslide susceptibility for every cell ($S_{LS[i]}$) against $\kappa_{LS[i]}$, for every cell i= 1,2,3,…N, where the size of the marker (the 'bubble') is proportional to the interquartile range of $S_{LS[i]}$. Results are only shown for Collazzone, as Oat Mountain S_{LS} takes discrete values, and thus is not easily visualised. In **Figure 6.12** we observe a 'wedge' shaped distribution: as median $S_{LS[i]}$ increases, the variability in κ_{LS} increases. That is, cells with higher median susceptibility may contain more landslide centroids (κ_{LS}) but also may contain no or few landslide centroids. It is also visually apparent that when median $S_{LS[i]}$ is greater than approximately 0.9, the size of the bubbles is smaller, meaning that there is little variability of S_{LS} within these cells. However, the values of κ_{LS} around the middle of the distribution ($0.4 \leq S_{LS} < 0.9$) are considerably 'noisier'.

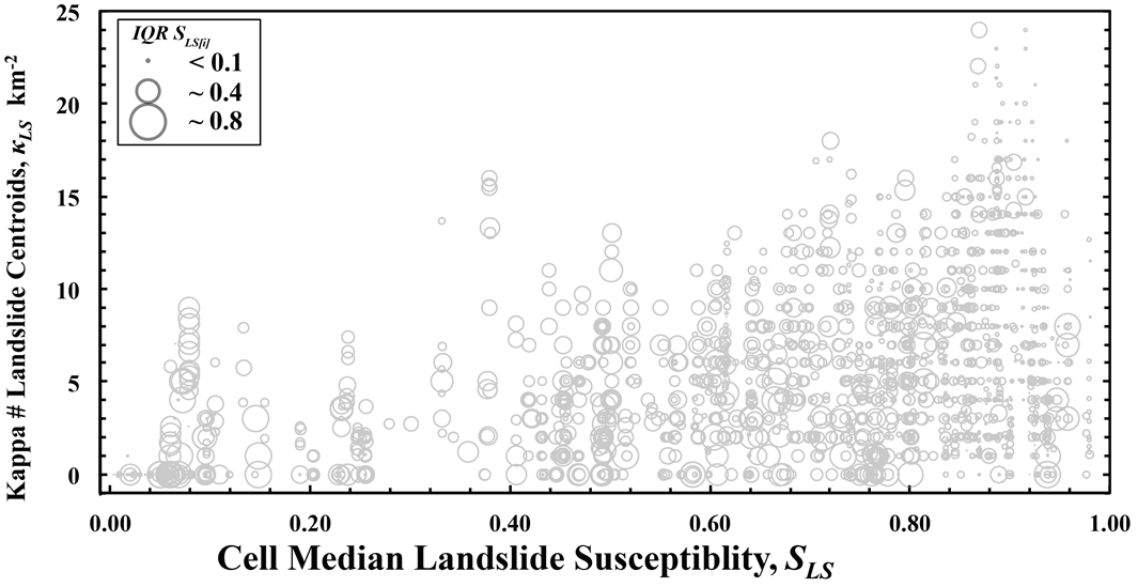


Figure 6.12 *Bubble plot of median cell landslide susceptibility ($S_{LS[i]}$) and Kappa-measure (κ_{LS}) of normalised number of landslide centroids km^{-2} where bubble size corresponds to the inter quartile range of landslide susceptibility in that cell (IQR $S_{LS[i]}$) for the Collazzone inventory. The Collazzone inventory has 413 landslide centroids over a 79 km^2 study region. A grid with a total of 181 (1 km × 1 km) cells is required to cover the entire study region. When using the moving window approach (cells moved 200 m at a time in both N and E directions, see* ***Section 6.2.2*** *for discussion) this equates to 2,231 (1 km × 1 km) cells. The Oat Mountain inventory is not shown, as S_{LS} is takes discrete values, and thus is not easily visualised.*

Figures 6.11 and **6.12** suggest a non-direct relationship between S_{LS} and κ_{LS}. i.e., higher landslide susceptibility does not necessarily result in a greater number of landslide centroids. In the case of the Collazzone landslide susceptibility map, low values of S_{LS} correspond to low probabilities of landslide occurrence, and high values correspond to high probabilities of landslide occurrence. However, middle values of S_{LS} correspond more to uncertainty in membership of either group (Rossi *et al.* 2010a), perhaps explaining the variability in κ_{LS} for middle values of landslide susceptibility (**Figure 6.12**). This has implications for the LRIM model (which will be fully described in **Chapter 7**), which uses a landslide susceptibility map to condition where more or less synthetically generated landslides are located throughout a study region (method summarised in **Section 6.1**). In the following **Section 6.4.2.3**, we compare the Kappa-measure for observed and LRIM generated triggered landslide event inventories.

6.4.2.3 Kappa-measure of Landslide Spatial Clustering in the LRIM Model

The Landslide Road Impact Model (LRIM) synthetically generates triggered landslide event inventory maps by randomly selecting landslide areas (A_L) and shapes (L/W) from statistical distributions described in **Chapters 4** and **5**. These landslides are then semi-randomly dropped over the region, conditioned by a susceptibility map, assuming that

more landslides will occur in high susceptibility regions and less in low susceptibility regions. The total number of landslides generated is defined by the average density (L_D) of landslides across the entire study area (N_{LS}/A_S km^{-2}). During LRIM development, we set L_D to equal the average density of the observed triggered landslide event inventories, and confronted the model output against the observed inventories to compare the spatial clustering of landslides. **Figure 6.13** shows some comparisons of the Kappa-measure for observed and modelled inventories for Collazzone and Northridge using (**A** and **B**) violin plots and (**C** and **D**) kernel density estimation. A violin plot is similar to a boxplot, showing the range of the dataset along the *y*-axis, but also the kernel density on the *x*-axis (Hintze and Nelson, 1998). Kernel density is essentially a 'smoothed' version of probability density (introduced in **Chapter 4**), showing where more or less of the observed data points lie (Rosenblatt, 1956).

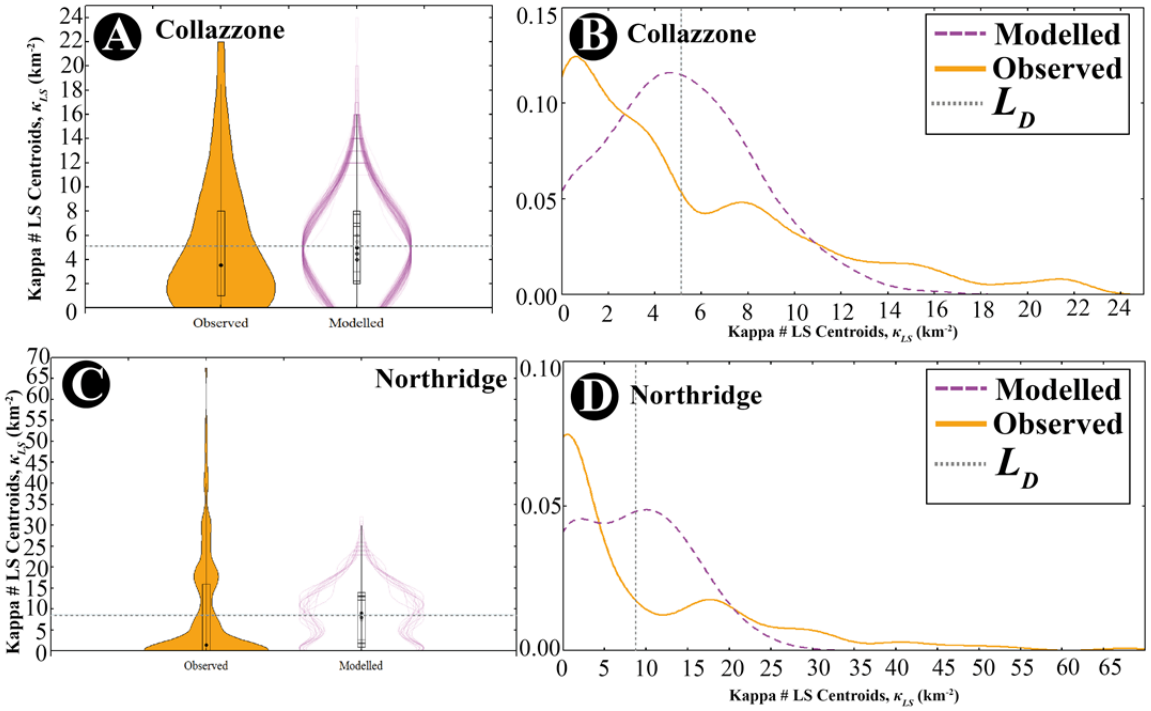


Figure 6.13 *Comparison of Kappa-measure number of landslide centroids km^{-2} between observed and modelled inventories for (**A** and **B**) Collazzone (subset of Umbria 1997 snowmelt inventory) and (**C** and **D**) Oat Mountain (subset of Northridge 1994 earthquake inventory). **Figures A** and **C** are violin plots, showing the distribution of the data as a boxplot on the y-axis, plus the probability density on the x-axis. **Figures C** and **D** are kernel density plots, which show the probability density distribution of the data (band width = 3 standard deviations) using a smoothing function. Modelled results are 10 iterations of the LRIM Model (briefly explained in **Section 6.1** and explained in detail in **Chapter 7**). The grey dashed line denotes the average landslide density (L_D) across the entire study area for Collazzone and Oat Mountain.*

Figure 6.13 illustrates the disparity between observed and modelled Kappa-measure for both locations, Collazzone and Oat Mountain. As shown in **Sections 6.4.2.2** and **6.4.2.3**, there is a high probability of observing a cell containing no landslide centroids in the

observed inventories. By the same means, a small proportion of cells contain a very large number of landslide centroids in the observed inventories, which averages out to the average density across the entire study area (L_D). Whereas, the modelled inventory Kappa-measure values tend to be closer to the average landslide density (L_D) across the entire region, and less of the data is in the tails of the distributions. This suggests that the spatial clustering of landslides occurs at a finer scale than can be modelled using a landslide susceptibility map. In the following **Section 6.5**, we summarise the findings from the pair-correlation and Kappa-measures and discuss the implications for the LRIM model.

6.5 Discussion and Implications for the Landslide Road Impact Model

In **Section 6.4.1**, we demonstrated using pair-correlation that landslide spatial clustering has some similarities across eight triggered landslide event inventories that vary in location, topography, landslide triggering mechanism and other factors. When normalised by the study area (A_{Circ}), although the probability density distributions of r were somewhat noisy, we were able to roughly separate the inventories into two types of clustering behaviour: (i) inventories dominated by 'low mobility' landslides (slips and slides) tend to cluster at a finer spatial scale and (ii) inventories dominated by 'high mobility' landslides tend to cluster at a coarser spatial scale (i.e., there are typically larger distances between landslides). Although not fully investigated, we believe the variability in probability density could be attributed to the structure of the landscape and the presence of very low landslide susceptibility zones (such as lakes, urban areas, etc.) across a region. In **Section 6.4.2** we showed using the Kappa-measure of normalised number of landslide centroids per square kilometre, that the relationship between landslide susceptibility (S_{LS}) and number of landslides that occur in a normalized (1 km × 1 km) cell (κ_{LS}) is not direct (i.e., high SLS does not necessarily result in high κ_{LS}) and noisy. By comparing some Landslide Road Impact Model (LRIM) outputs to observed Kappa-measure values (**Section 6.4.2.3**), we found that we do not recreate spatial clustering of triggered landslides in the LRIM at the scale seen in the observed inventories. **Figure 6.14** reinforces this idea, showing the cumulative landslide area (normalised by the sum of all individual landslide polygon areas (A_L) in an inventory) with the value of susceptibility at each landslide centroid for both the observed Collazzone inventory and 100 synthetic triggered landslide event inventory maps generated in the LRIM model.

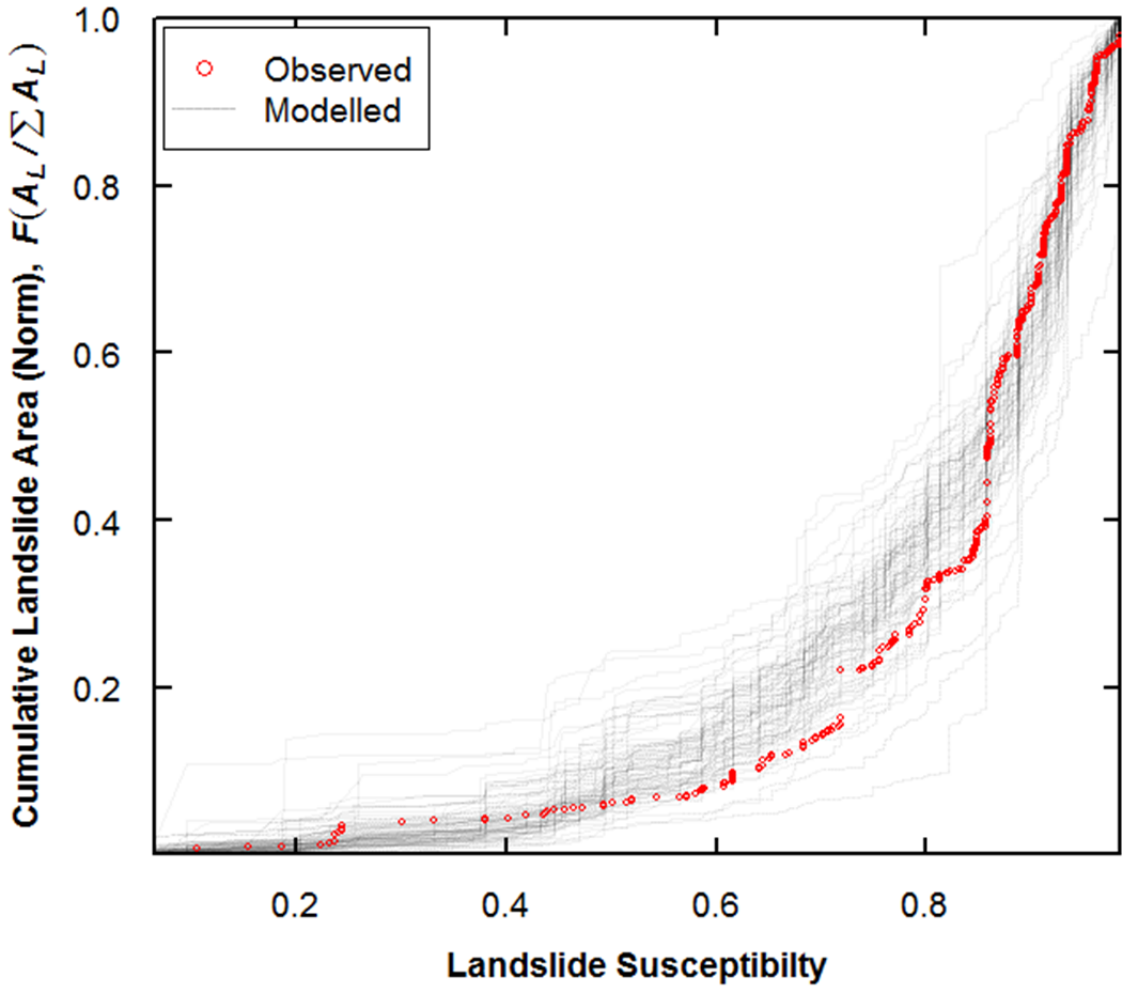


Figure 6.14 *Cumulative area of individual landslide polygon areas (A_L) divided by total landslide area for that inventory, and plotted against landslide susceptibility value (S_{LS}) at the location of that landslide. Red dots denote the observed landslide inventory of 413 landslides triggered by snowmelt in Collazzone in 1997. Each grey line denotes 1 of 100 synthetic triggered landslide event inventories created in the Landslide Road Impact Model (LRIM) to match the overall landslide density (L_D) of the observed inventory (explained in detail in the following Chapter 7).*

In **Figure 6.14**, we observe some disparity in the cumulative distribution of landslide area between the modelled and observed inventories, particularly for the middle values of landslide susceptibility ($0.4 \leq S_{LS} < 0.8$). In the observed inventory, < 20% of the total landslide area (ΣA_L) of the inventory occurs where susceptibility is $S_{LS} \leq 0.7$. After this point ($S_{LS} > 0.7$), the gradient of observed normalised cumulative landslide area increases. Whereas, the LRIM normalised cumulative landslide area increases more steadily from low to high values of susceptibility (S_{LS}), meaning that we model more landslide areas in medium susceptibility zones than is observed.

More in depth confrontation of the final LRIM model is presented in **Chapter 7**, but research done in this **Chapter 6** was performed during the development stages of LRIM. We wanted to assess whether the spatial clustering of landslides has an impact on the number of road blockages simulated by LRIM (and thus whether it would be useful to

incorporate clustering into the model). To trial the introduction of spatial clustering of landslide centroids into the LRIM for the Collazzone test site, we used the inventory of landslide centroids triggered by 1997 snowmelt in Umbria (**Table 6.1**), of which the Collazzone inventory is a subset (a 79 km^2 basin within the Umbria region). We then applied the following steps (illustrated in **Figure 6.15**):

1. *Plot the observed landslide centroids* across the Umbria region on a map (**Figure 6.15A**).
2. *Plot the 'actual' location of the Collazzone study region* within the Umbria region (**Figure 6.15B**).
3. *Randomly translate the Collazzone study region extent* to another (*N, E*) location within the Umbria study region (**Figure 6.15C**).
4. *Select all landslide centroids within the 'new' study region.*
5. *Transform these landslide centroids* so they are in the study region (**Figure 6.15D**).
6. *Add area and shape to each landslide centroid* as per the standard LRIM configuration, outlined in **Chapter 7**.
7. *Overlay synthetic landslide inventory with map of road network* to identify number of roads blocked by landslides (N_{BI}).
8. *Repeat Steps (3-7)* ten times to generate ten synthetic triggered landslide event inventory maps.
9. *Identify road blockages*. This is done by overlaying each synthetic landslide inventory generated in Steps (3–7) with a road network map of the region. Landslides within 50 m and 100 m of the road are also identified by overlaying the synthetic landslide inventories with a road network map where the lines have been buffered by 50 m and 100 m either side of the road line.
10. *Compare the observed number of road blockages to the modelled number of road blockages* (**Figure 6.16**).

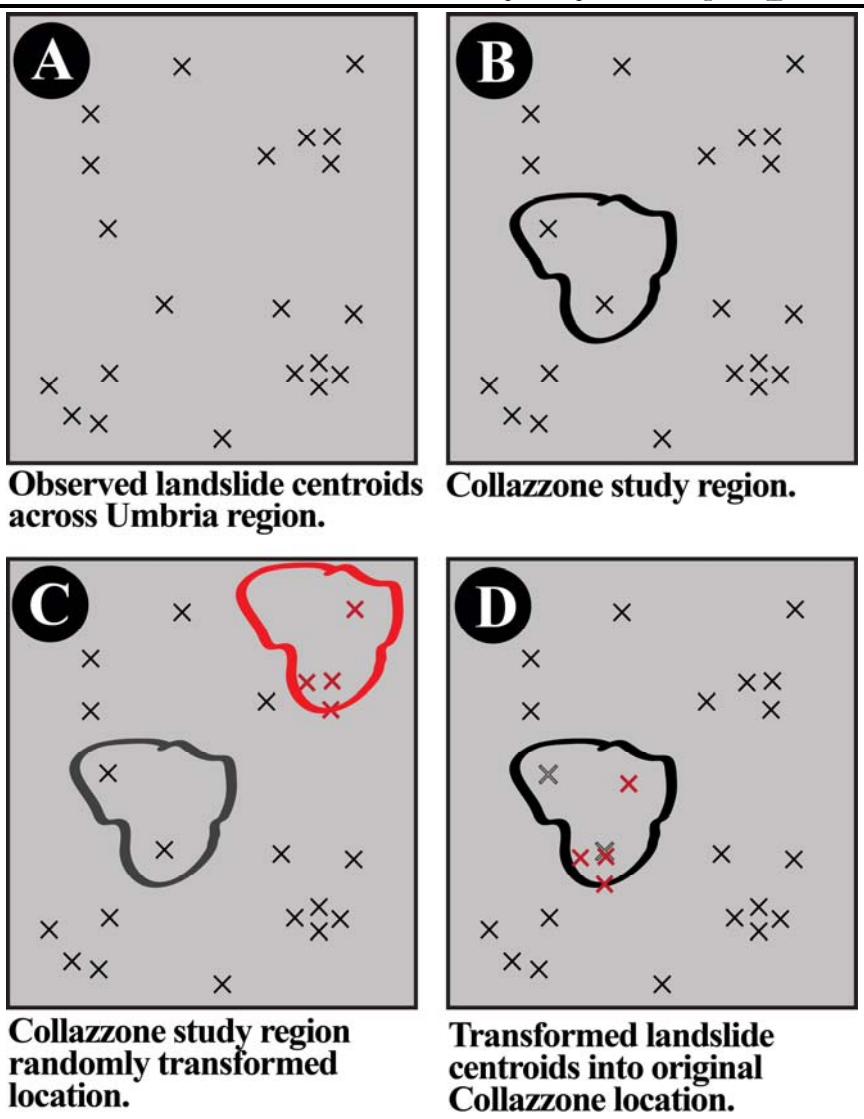


Figure 6.15 *Illustrative example of a quick method used to trial the introduction of observed spatial patterns of landslide clustering into the Landslide Road Impact Model (LRIM) for the Collazzone region.* ***(A)*** *Map of observed landslide centroids across the Umbria region.* ***(B)*** *Map of the actual location of the Collazzone study region within the Umbria region.* ***(C)*** *The map of the Collazzone study region extent is randomly translated by selecting a new set of (N, E) coordinates)* ***(D)*** *The landslide centroids occurring within this transformed Collazzone study region are transformed back to the true Collazzone location, and true landslide centroids within the Collazzone region deleted. This creates a synthetic map of landslide centroids that exhibit clustering.*

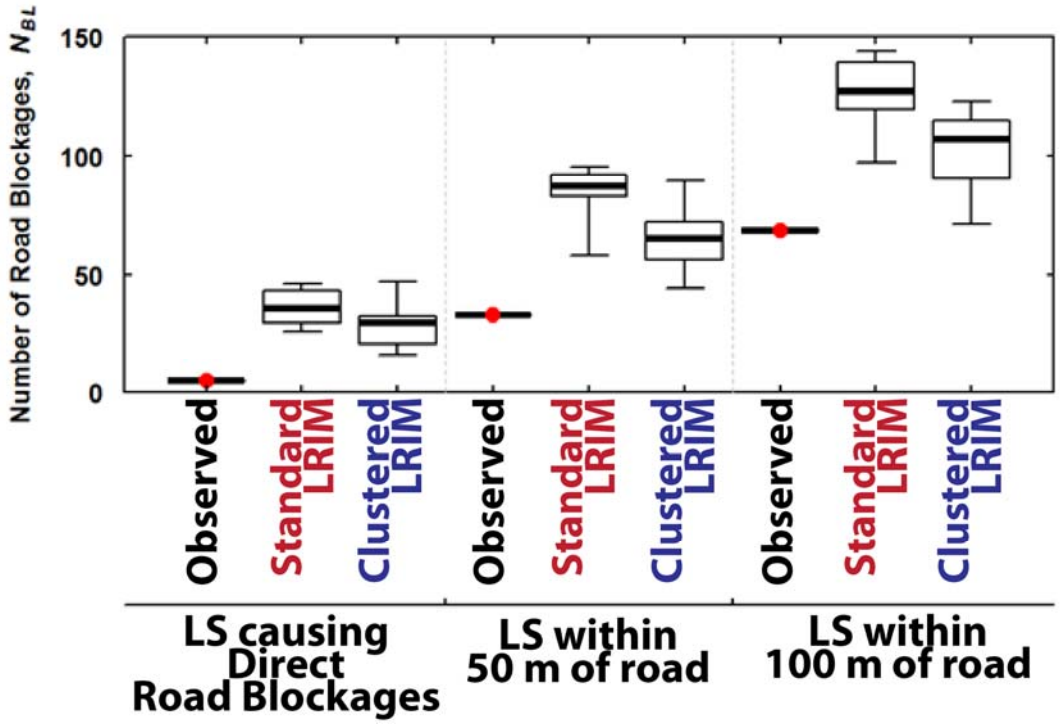


Figure 6.16 *Comparison of observed number of road blockages (N_{BL}) to the 'standard' Landslide Road Impact Model configuration (Standard LRIM) where no clustering is introduced, and LRIM adapted to incorporate spatial clustering of landslides (Clustered LRIM). In each case, LRIM is run ten times. Observed road blockages are those identified by overlaying the observed triggered landslide event inventory of 413 landslides triggered by 1997 snowmelt in the Collazzone region with a road network map for the region. Both 'Standard LRIM and 'Clustered LRIM' are configured so that landslide density (L_D) is the same as the observed inventory. Road blockages are measured as landslide polygons directly intersecting with the road line, landslides that occur within 50 m of the road and landslides that occur within 100 m of the road.*

Figure 6.16 shows boxplots of the observed number of road blockages for the Collazzone region, calculated by overlaying the Collazzone 1997 snowmelt triggered landslide event inventory map with a 153 km road network map for the region. Also calculated are the number of landslides within 50 m and 100 m either side of the road, done by buffering the road map by 50 m and 100 m either side of the road line. The LRIM model was run for ten iterations, both with standard and 'clustered' configurations (outlined in Steps 1–10 above), and the number of road blockages (and landslides within 50 m and 100 m of the road) counted for each iteration. Modelled number of road blockages over ten LRIM iterations is shown as boxplots in **Figure 6.16**. **Figure 6.16** illustrates that incorporating spatial clustering of landslides into LRIM does not significantly improve the model skill in predicting the number of observed road blockages, or landslides in close proximity to the road. This may be due to the semi-random Monte Carlo approach used (in some LRIM simulated inventory maps, landslides may cluster away from roads, and in other LRIM simulated inventory maps, landslides may cluster close to roads, resulting in a similar overall range of number of road blockages in **Figure 6.16** to the standard LRIM

configuration. Because the incorporation of landslide clustering in LRIM did not significantly improve model skill, and would likely increase the time to run each LRIM iteration, we decided not to include this into the LRIM model.

6.6 Summary and Conclusions

We have illustrated the application of two methods (pair-correlation and Kappa-measure) to measure spatial clustering of landslides, showing that landslides tend to cluster at a small spatial scale, supporting the findings of Guzzetti *et al.* (2006a) and Zaitchik *et al.* (2003). We also suggest that the spatial clustering of triggered landslide event inventories may be somewhat semi- 'generally applicable' when normalised by the extent of the affected region, which is in somewhat of a contrast to other research looking at the relationship between the landslide trigger and the resulting spatial pattern of landslides (e.g., Meunier *et al.*, 2013). Through the analysis of the relationship between the Kappa-measure of normalised number of landslides per square kilometre grid cell and median landslide susceptibility within that cell, we have highlighted the (already known) uncertainties of using landslide susceptibility maps to forecast the spatial occurrence of landslides (e.g., Guzzetti *et al.* 2006a): that some high landslide susceptibility zones may fail multiple times in one triggering event, others may not fail at all, and some lower susceptibility zones may also fail. However, the methods shown here (e.g., **Figure 6.11**) may aid others in forecasting what proportion of high landslide susceptibility zones may encounter one or more landslides. The research presented in this **Chapter 6** was somewhat of a 'means to an end' to investigate what effect spatial clustering of landslides may have on number of road blockages, and whether this is an important variable for the LRIM model. As it was shown to not significantly improve model skill, we do not use this research in the following **Chapter 7**. However, we believe the methods shown here could be applied to additional triggered landslide event inventories and also investigated further.

PART III: DEVELOPMENT OF A LANDSLIDE-ROAD IMPACT MODEL

Chapter 7 The Landslide-Road Impact Model (LRIM)

Chapter 7 Summary

This primary research chapter outlines the development and application of a Landslide-Road Impact Model (LRIM) to simulate triggered landslide events and their impacts upon road networks. This model exploits the statistical behaviour of triggered landslide event inventories established in **Section 2** of this thesis, and after application in study regions is then confronted with 'real' landslide inventories. The LRIM creates 'synthetic' triggered landslide event inventories by randomly sampling landslide areas and shapes from already established statistical distributions (described in **Chapter 4** and established in **Chapter 5**, respectively). In the methodology, these landslides are then semi-randomly dropped across a study region, conditioned by a landslide susceptibility map. The resulting synthetic triggered landslide event inventory is overlaid with a road network map and the number, size, location and network impact of road blockages calculated. This process is repeated hundreds of times in a Monte Carlo type simulation. Because the statistical distributions and approaches used in the model are thought to be generally applicable for low-mobility triggered landslides in many medium- to high-topography regions throughout the world, relatively little detailed local data is required to run the model. Coupled with an open-source modelling approach, this model may be applied to many regions where triggered landslide events occur. We apply LRIM, and confront with observed data for three study regions: (i) Collazzone (Central Italy) where rapid snowmelt triggered 413 landslides in January 1997; (ii) Oat Mountain (Northridge, USA), where the Northridge Earthquake triggered 1,356 landslides in January 1994; (iii) Su-Hua (northeast Taiwan) where Typhoon Megi triggered 149 landslides in 2010. We find that when the landslide susceptibility map is adjusted along road corridors to take into account interactions between landslides and roads, the model matches reasonably well the observed results. In Collazzone (length of road = 153 km, landslide density = 5.2 landslides km^{-2}), the median over 100 model runs was 5 (±2.5 S.D.) road blocks, compared to the observed number of 5 road blocks. In Oat Mountain (length of road = 780 km, landslide density = 8.7 landslides km^{-2}), the median over 100 model runs was 108 (±17.2 S.D.) road blocks compared to the observed number of 48 road blocks. In Su-Hua (length of road = 265 km, landslide density = 0.35 landslides km^{-2}), the median over 100 model runs was 8 (±3.8 S.D.) road blocks compared to the observed number of 18 road blocks. This Landslide-Road Impact Model presents an open-source low-data methodology for the stochastic simulation of potential simultaneous road network impacts.

*List of abbreviations used in **Chapter 7***

Abbreviation	Meaning	Section Introduced
DEM	Digital Elevation Model	7.3
GIS	Geographic Information System	7.2
LRIM	Landslide-Road Impact Model	7.1
LS	Landslide	7.3
MLE	Maximum Likelihood Estimation	7.3

Abbreviation	Meaning	Section Introduced
pdf	Probability density function	7.3
SD	Standard deviation	7.1

List of variables used in ***Chapter 7***

Variable	Meaning	Section Introduced
α	Alpha measure of network connectivity	7.5
Γ	Gamma function ($\Gamma(x) = (x-1)!$) Where ! = factorial	7.3
$\varkappa_{AR}$	Normalised area of road network that has been buffered by 100 m either side of the road line, per square kilometre	7.5
$\varkappa_{BL}$	Normalised number of landslides within 100 m of the road, per square kilometre	7.5
$\varkappa_{LS}$	Normalised number of landslide centroids per square kilometre	7.5
ρ	Scale parameter for inverse Gamma pdf	7.3
ϕ	Soil friction angle	7.4
a	Shape parameter for inverse Gamma probability density function	7.3
A_L	Individual landslide (L) area	7.1
c	Soil cohesion	7.4
D	Distance from landslide centroid to nearest road	7.3
D_R	Damage to road (R)	7.2
E	Number of edges in a network	7.5
$F(x)$	Cumulative distribution function of x	7.3
g	Soil density	7.4
k	Hydraulic conductivity	7.4
L	Length	7.2
L/W	Landslide length-to-width ratio	7.1
L_D	Average landslide density across a region	7.3
L_{RD}	Length of road (RD)	7.3
M_W	Earthquake moment magnitude	7.1
n	Number or count (different variable types)	7.3
N	Number of nodes in a network	7.5
(N, E)	Northing and Easting Coordinates	7.3
N_{BL}	Number of road blockages	7.5

Variable	Meaning	Section Introduced
N_{LS}	Total number of landslides (LS)	7.3
p	Number of isolated subgraphs in a network	7.5
s	Location parameter for inverse Gamma pdf	7.3
S_{LS}	Landslide susceptibility	7.3
x	Random value selected from uniform distribution	7.3
z	Vertical thickness of soil	7.4

List of equations used in ***Chapter 7***

Equation Number	Equation	Section Introduced
7.1	Inverse-Gamma probability density function	7.3
7.2	Alpha measure of road network connectivity	7.5

7.1 Introduction

In many medium- to high-relief topographic regions of the world, triggered landslide events pose a risk to infrastructures, livelihoods and assets. A triggered landslide event includes all of the landslides resulting from one triggering event such as an earthquake or heavy rainfall. Depending on the location and trigger, a triggered landslide event might include just several landslides to hundreds of thousands of landslides, occurring across the region within a short space of time (e.g., minutes to days). A number of roads can become blocked by landslides in a given triggered landslide event, which can make it difficult to move about a region by road and conduct rescue and recovery operations. In extreme cases, people may become entirely isolated for days to weeks. A recent example (Murty *et al.*, 2012; Martha *et al.,* 2014) is of remote villages in Sikkim, India that were isolated for three weeks following the 2011 (M_W = 6.9) Sikkim Earthquake, which triggered > 1,196 landslides and reactivated other landslides. Based on a survey of twelve stretches of road across the affected area, Chakrabotrty *et al.* (2011) observed 210 landslides in close proximity to the road. Murty *et al.* (2012) noted that at least three settlements were cut off for over three weeks because of these road blockages. Murty *et al.* (2012) further reported that the earthquake was proceeded by a week of heavy storms, making it difficult to reach people by helicopter and increasing the magnitude of the impact caused by triggered landslide event. They concluded that the damage and physical isolation caused by the

landslides compounded the impact of the earthquake, which damaged other critical lifelines such as government buildings, schools and hospitals.

There are many other examples (discussed in **Section 7.2.2**) where landslides impact road networks or other types of critical infrastructure (e.g., pipelines (Cevik *et al.*, 2003); railways (Bednarik *et al.*, 2010); power networks (Petrova, 2011), etc.). Although our understanding of the risk and impact of an individual landslide blocking a road is reasonably well developed (e.g., Hearn *et al.,* 2011; Remondo *et al.,* 2008; Jaiswal *et al.*, 2011; Voumard *et al.*, 2014), we will show in **Section 7.2** of this chapter that modelling of simultaneous road blockages caused by landslides and the resulting network wide impact has not received much attention in the literature. **Section 7.2** further shows that this type of network wide analysis is more advanced for modelling other hazards such as earthquakes, and that some of these techniques could be applied to triggered landslide events, which we do in the development of the Landslide-Road Impact Model (LRIM) introduced in **Section 7.3**. In **Section 7.3**, we present research on developing a computer model that simulates a triggered landslide event inventory map (see **Chapter 2** for a background on landslide inventories) by randomly selecting landslide areas (A_L) and ellipse length-to-width ratios (L/W) from 'generally applicable' probability density functions discussed in **Chapter 4** (triggered event landslide area statistics) and **Chapter 5** (triggered event landslide length-to-width statistics) We then take these synthetic landslide ellipses that are drawn from known probability distributions, and semi-randomly 'drop' these on regional and local road networks, with the impact of those landslides on the road networks calculated probabilistically. **Section 7.4** introduces the three sites where LRIM is applied, and **Section 7.5** presents results for each site. **Section 7.6** discusses results and concludes the chapter. We believe this methodology could have general applications to the simulation of potential impact for landslide triggered events.

The Landslide-Road Impact Model (LRIM) developed here as part of this PhD, also contributed to Deliverable 6.4 of the two-year (3/2013−2/2015) EU FP7 LAMPRE (Landslide Modelling and Tools for Vulnerability Assessment Preparedness and Recovery Management) research project, and was a collaborative effort led by myself (Faith Taylor, KCL) with contributions from Bruce D. Malamud (KCL) in his role as PhD supervisor and PI for KCL on the LAMPRE grant, and three scientists from CNR-IRPI (Italy) also involved with the LAMPRE project: Michele Santangelo who collaborated in the development of the GRASS-GIS Python codes for the 'standard' version of the model

(outlined in **Section 7.3.2.6**); Ivan Marchesini who helped develop initial versions of some of the code used here; and Fausto Guzzetti in half a dozen discussions.

A summary of the sections in **Chapter 7** is the following:

7.2 Review past and current research that examines (i) understanding the impact of landslides on road networks and (ii) modelling other hazard impacts on road networks.
7.3 Outline our Landslide-Road Impact Model (LRIM) in terms of input data, model methodology and types of output produced.
7.4 Introduce three sites where the LRIM has been applied.
7.5 Present results for the three sites where the LRIM has been applied and confronted with real triggered event landslide events that occurred in those regions.
7.6 Discuss potential future applications of the model and summarise our findings.

7.2 Current Research on Landslide and Other Hazard Impacts on Roads

7.2.1 What is Landslide Road 'Impact'?

In this **Section 7.2.1** we define how we use the term 'impact' and review different types of landslide road impact. Here, we use the term 'impact' to refer to the consequences of a landslide intersecting with a road. This could entail (but is not limited to):

(i) *Direct consequences* such as damage to the road surface caused by the landslide.
(ii) *Relatively tangible indirect consequences* such as additional cost of fuel used to drive the detour route.
(iii) *Less tangible indirect consequences* such as loss of business resulting from difficulty in accessing a location.

The spectrum of direct and indirect impacts caused by landslides on road networks is wide, and conceptualised differently by different research groups (as we will show in this **Section 7.2.1**). Indeed, in a recent review of the literature, Ardizzone *et al.* (2014b) showed that out of 75 articles investigating landslide impact on structures and infrastructures, there was almost an even spread between those conceptualising the problem as:

(i) *Quantifying the economic cost* of direct damages and impacts.
(ii) *Quantifying the physical impact* of number and location of landslides blocking roads.
(iii) *Calculating the risk of road impact,* defined as the probability of a landslide of a given magnitude occurring on the road network and the consequences of this.

(iv) *Calculating the vulnerability to impact* e.g., by linking the size of a landslide to the amount of damage it causes.

In **Table 7.1** we outline the typical types of impact caused by landslides on road networks, based upon the following sources: (i) literature reviewed throughout this thesis, (ii) the work performed on the National Landslide Database of Great Britain (**Chapter 3**), and (iii) more generally following media reports of landslide road impact over the past four years. Consequently, we believe this list of landslide-road impact types is reasonably representative and covers the main spectrum of impacts encountered. In **Table 7.1**, we also show different ways these impacts would be categorised based on three different impact categorisation schemes found in the peer review and grey literature.

Table 7.1 *List of 17 main types of impact caused by landslides on road networks (column 2). These impacts are classified into three ways of classifying impact that we found in the literature (Ardizzone et al., 2014b; Mooney (personal communication, 2014) and Glade and Crozier 2005).*

		Impact Classification Scheme									
		Ardizzone *et al.* (2014)			**Mooney (2014)**			**Glade and Crozier (2005)**			
		Physical	**Economic**	**Social**	**Damage**	**Denial**	**Danger**	**Direct**	**Indirect**	**Acute**	**Chronic**
Impact Category	**Type of Impact** / **Definition of Impact**	Physical inter-section of landslide and road	e.g., loss of work	e.g., Injury, fatalities, migration	Physical damage	Loss of access to location/ facility	Immediate injury/ death	Incurred by direct physical contact with the landslide	Changes in property/ behav-iour of other system	Short lived	Manifest over a longer period
Physical	*Landslide landing on a road*	×			×			×		×	
	Landslide damaging a road surface	×			×			×		×	×
	Landslide destroying a road surface	×			×			×		×	
	Landslide hitting a vehicle	×		×	×		×	×		×	
Route Connectivity	*Landslide partiatlly blocking a road*	×				×		×		×	×
	Landslide fully blocking a road	×				×		×		×	×
	Landslide temporarily closing a road		×	×		×		×		×	
	Landslide permanently closing a road		×	×		×		×			×
People	*Road blockage causing a delay*		×	×		×			×	×	×
	Road blockage causing a detour		×	×		×			×	×	×
	Road blockages trapping vehicles	×	×	×	×	×			×	×	
	Road blockage resulting in road isolation of settlements	×	×	×		×			×	×	×
	Road blockage preventing evacuation	×	×	×		×			×	×	
Damage/ Repair	*Landslide road damage requiring repair*		×		×				×	×	×
	Landslide road damage requiring complete rebuilding		×		×				×		×
	Landslide road damage requiring a new route		×		×				×		×
Other	*Landslide triggering secondary hazard with road impact*	×			×				×	×	×

Table 7.1 illustrates five landslide-road impact categories (physical, route connectivity, people, damage and other), which account for 17 different impact types. Of these, 9 are indirect impacts, and can affect a much wider spatial extent than the actual landslide. For instance, a very small area landslide (e.g., A_L < 20m^2) has the potential to cause national scale disruption and large economic losses if it intersects with an important road with few alternative routes (e.g., Winter (2015) gave the example of a small debris flow that blocked L <10 m length (L) of the A83 road in Scotland in 2007 but due to few alternative routes, the area of the affected region was estimated to be > 2,500 km^2). Other than a landslide directly hitting a vehicle (potentially causing a fatality), most of the landslide-road impacts can be classified as having economic or social consequences. Examples include the cost of

repairing a road or the social impacts of isolation from other population centres (Damm and Klose, 2015). The review of current literature looking at modelling the impact of landslides on road networks (**Section 7.2.2**) will illustrate that our current understanding of landslide-road impact is limited by poor recording of past events and the large efforts required to collect regionally specific data. Whereas, **Section 7.2.3**, shows that in other hazard modelling disciplines, there are approaches to modelling impact that sometimes circumvent the issue of data paucity, which it may be possible to apply to the problem of modelling triggered landslide event road network impact in LRIM (introduced in **Section 7.3**).

7.2.2 Current Approaches to Modelling Landslide-Road Impact

This **Section 7.2.2** reviews current approaches to understanding and modelling landslide road impact found in the literature. These are broadly classified as:

7.2.2.1 *Forensic analysis of cost*

7.2.2.2 *Economic impact modelling*

7.2.2.3 *Spatial intersection of landslide inventories and road network maps*

7.2.2.4 *Investigating the link between roads and landslide susceptibility*

7.2.2.5 *Traffic impact modelling*

7.2.2.1 Forensic Analysis of Cost

Much of our understanding of landslide-road impact and policy making comes from forensic reconstruction of the cost of previous landslide events, although this method is subject to uncertainties and issues with data recording. Direct cost of road repair or rebuilding can often be found through contact with highways agencies (e.g., De Graff *et al.*, 1989 for Tobago; Crovelli and Coe, 2009 for the state of California, USA; Remondo, 2008 for a region in northwest Spain; Jaiswal *et al.*, 2010a for a region in Southern India), or sometimes be recorded as part of a landslide database (e.g., Klose *et al.*, 2014 for a region in northwest Germany). A few examples exist in the literature of case studies where following a triggered landslide event, a systematic survey of damages is performed soon after the event (e.g., Ardizzone *et al.*, 2014b for the Marche region in Central Italy, where > 1,500 landslides were triggered by heavy rainfall in winter 2013).

Broadly speaking, the cost of landslides estimated through forensic reconstruction is generally incomplete due to issues such as:

(i) *Landslide damage is not distinguished from damage caused by the triggering event* (e.g., landslide damage is recorded as earthquake damage) (Kirschbaum *et al.*, 2010).

(ii) *Both public and private insurance may exclude damage caused by landslides so records of loss do not exist* (Willis *et al.* 2014).

(iii) *Information on damage to the infrastructure may be held and recorded by different agencies, each with different systems*. For example, Willis *et al.* (2014) note that in California, Caltrans central office keep a database of major "emergency opening" projects that cost > US $250,000 for state highways; Caltrans district offices are responsible for keeping paper records of more minor emergency opening projects; whereas the California Department of Transport keep records of long term projects to reopen or rebuild highways.

Arguably, reconstructing the cost of previous events is the most established method of measuring landslide road impact, although cost is almost always underestimated due to difficulties of data collection, and is somewhat difficult to extrapolate from the magnitude of an individual past event to different magnitude landslide events in the future (Wills *et al.*, 2014). Later in **Section 7.2.2.3** a study that has combined this type of forensic reconstruction of cost with a more spatial landslide inventory-based approach will be described.

7.2.2.2 Economic Impact Modelling

Section 7.2.2.1 showed how we can record the cost of past damages. However, this is not particularly useful in forecasting future events, unless the future event follows the same 'template' as a past event (e.g., same magnitude and location). Thus economic modelling approaches aim to extrapolate/interpolate past observations of cost to establish relationships between the magnitude or spatial extent of a landslide event and the economic impact, which can then be scaled up or down to model future events.

Successful examples of landslide economic impact modelling tend to be in the developed world where more resources are available for data recording, and there is a higher aversion to risk (Klose *et al.*, 2014), due to the lack of observational data of past events, discussed in **Section 7.2.2.1**. Examples of fairly 'comprehensive' modelling include Wills *et al.* (2014), who investigated the impact of rainfall triggered landslide events on the transport and built

infrastructure of California. They used observations of infrastructure damage from previous winter storms occurring in 1969 and 1986 recorded by Caltrans and other public/private bodies responsible for infrastructure maintenance, agglomerated to census tract level. Landslide susceptibility was also calculated at census tract level and the relationship between susceptibility and losses correlated. The analysis of different storms of different intensity and regional variations within a single storm allowed some extrapolation of the link between rainfall intensity and amount of economic loss caused by landslides, shown in **Figure 7.1**. Wills *et al.* (2014) is one of the few studies identified that looks at triggered landslide events (i.e., multiple landslides occurring in a short space of time at the result of a given trigger), although it focuses on direct damage to the road infrastructure rather than the indirect impact of simultaneous road blockages on e.g., traffic flow. Klose *et al.* (2014) also performed cost extrapolation from a database of local landslide-road impact costs to the regional level, to estimate the annual cost of landslide damage to infrastructure per kilometre of highway in northwest Germany.

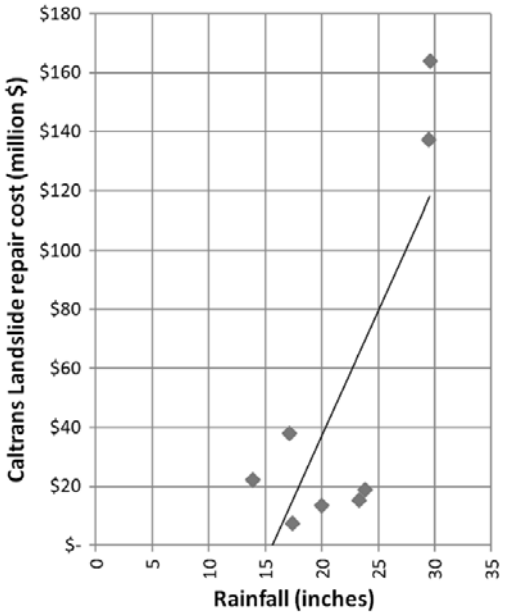


Figure 7.1 *Example of economic impact modelling where observed rainfall amount is correlated with reported cost of repairs to road networks caused by landslides for the state of California, USA. Figure from Wills et al. (2014).*

In a low-income country setting, Hearn *et al.* (2007) performed economic impact analysis of landslides on 1,500 km of the road network of Lao PDR. Relatively simple estimations of the direct cost to remediate previous landslides impacting the road were calculated based on general values. These were compared to the cost of proactive measures to prevent landslide occurrence. In a cost-benefit analysis, they found that the benefits of introducing proactive measures were marginal. However, when the indirect cost of loss of productivity

(estimated from a simple relationship between average daily traffic flow and cost per hour per vehicle queuing due to landslide disruption), that proactive measures became considerably more cost effective.

7.2.2.3 Geographic Information System (GIS) Use of Inventories to Calculate Physical Impact

The examples described in the previous **Section 7.2.2.2** focused on the economic impacts of an event, and were less focussed on the spatial occurrence or physical impact of landslides on road networks. **Section 7.2.2.3** discusses studies that overlay landslide inventories with road network maps to understand the spatial occurrence of landslide-road network impact (example in **Figure 7.2**). This is done because it is often assumed that the majority of landslides that will occur in the future will occur in relatively close proximity to where landslides have occurred in the past (Guzzetti *et al.*, 2003).

For example, Guzzetti *et al.* (2003) overlaid three different types of landslide inventory map (inventory types outlined in **Chapter 2**) with a regional road network map to quantify landslide-road interactions in the Umbria region, Central Italy. The inventories used were: two triggered landslide event inventory maps (4,235 landslides triggered by snowmelt in 1997; 220 rockfalls triggered by an earthquake in Sept-Oct 1997), one seasonal inventory map (1,072 landslides triggered by rainfall 1937-1941) and one geomorphological inventory map (47,414 landslides). A line map of the regional road network was buffered by various widths corresponding to the road hierarchy (20–150 m buffer widths) to account for (i) the paved area of the road, (ii) the fact that a landslide that occurs in close proximity to the road may result in some road network impact and (iii) positional mismatches between the datasets.

Each inventory was overlaid with the buffered road network map and the number of road blockages, number of landslides per kilometre of road and percentage of landslide inventory intersecting with the road network quantified. Summary descriptive statistics like this are easily calculable in a GIS and allow the comparison of different scenarios (e.g., different numbers of landslides or length of road network). Although the snowmelt triggered landslide event inventory and the seasonal rainfall inventory differ in total number of landslides, the percentage of the inventory intersecting with the buffered road network was similar in both cases (between 2.5–2.7 % of landslides in each inventory). In a second paper using the same inventories and methods, Reichenbach *et al.* (2002) found that the triggered rockfall event caused more road network impact (3.8 landslides per 10 km of

road) than the snowmelt triggered inventory (1 landslides per 10 km of road), which they attributed to the earthquake triggered inventory being dominated by high mobility rockfalls interacting with roads in valley floors. The average number of kilometres of road between landslides was generally higher for highways than local roads (i.e., less landslides intersect with highways than local roads, most likely due to the stricter engineering and design criteria to build a highway (FAO, 1998)). In a more in-depth study of the same region, Galli and Guzzetti (2007) investigated the relationship between landslide area (A_L) and damage to (i) paved roads and (ii) secondary, mainly unpaved roads for 73 landslides that intersected the road network in the period 1982 – 2005. Damage was ranked heuristically between $D_R = 0$ (no damage) and $D_R = 1$ (complete destruction) based on how much of the road was blocked by the landslide, impact on the local population, time taken to remediate and engineering characteristics of the road. Damage values $D_R < 0.2$ correspond to 'aesthetic' damage, $0.2 < D_R < 0.6$ 'functional' damage and $D_R > 0.6$ 'structural' damage. From this, landslide vulnerability thresholds were calculated, showing a positive relationship between landslide area (A_L) and the level of damage to the road network.

Ardizzone *et al.* (2014) built upon Guzzetti *et al.* (2003)'s approach, and combined this with economic modelling (**Section 7.2.2.2**). Shortly after a rainfall triggered landslide event in the Marche Region, Central Italy in 2013, an inventory of landslides was created. This was overlaid with the regional road network map to identify where landslides had intersected the road network, and what lengths of the road were blocked (example of GIS overlay shown in **Figure 7.2**). The cost of each repair was then investigated by liaising with the local government responsible for the repairs caused by landslides. This allowed the calculation of average cost of repairs per meter of road blocked and cost per square meter of landslide in a region. Because the analysis was performed soon after the event and used a substantially complete inventory of the triggered landslides, the authors could be confident that they obtained a relatively complete measure of the direct cost of road blockages caused by landslides in a triggered landslide event. This combination of spatial physical impact recording and forensic analysis of cost can be considered a 'higher confidence' estimate of impact because there is a high degree of confidence that all road blockages have been recorded (through the use of a landslide inventory).

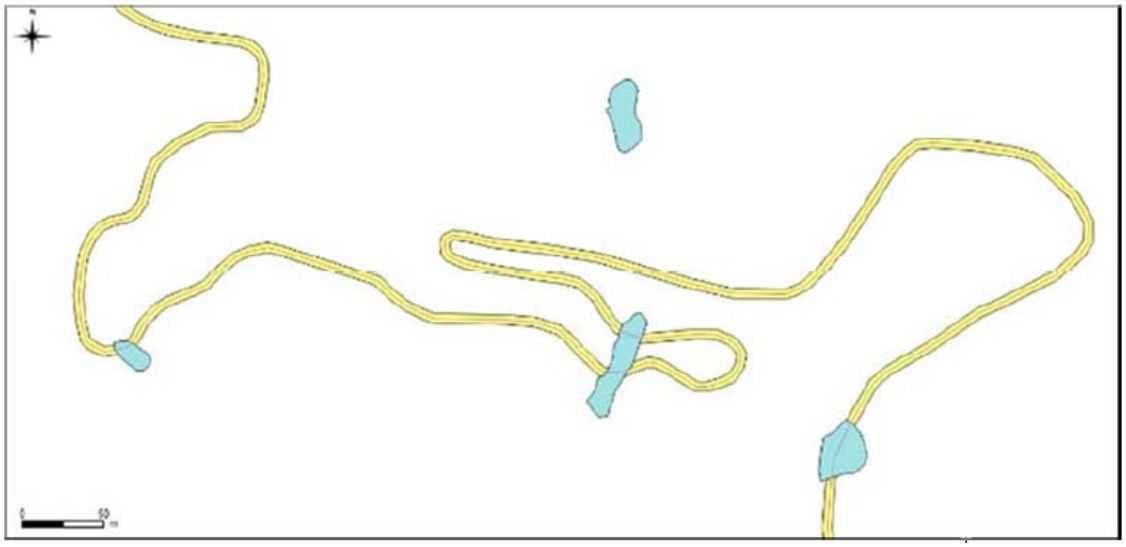


Figure 7.2 *Example of a GIS overlay of triggered landslide event inventory (blue polygons) and buffered road network (yellow lines) for landslides triggered by heavy rainfall in 2013 in the Marche Region, Central Italy. Three landslides have intersected with the road. One of these landslides has intersected the road twice. Figure from Ardizzone et al. (2014b).*

7.2.2.4 Roads and Landslide Susceptibility

Some of the studies described in **Section 7.2.2.3** using GIS overlays of landslide inventories and road network maps noted that landslide incidence may be increased near to roads due to factors such as (i) where roads are built relative to what portions of the slope are most susceptible to failure and (ii) road building techniques such as slope modification and drainage. This **Section 7.2.2.4** describes studies which have looked in more detail at the interactions between roads and landslide occurrence. First, studies that have looked at the density of landslides within buffer zones of a road are introduced. Second, studies that have built upon this work to calculate landslide susceptibility maps that take into account road presence are described. Third, studies that use landslide susceptibility maps (which may or may not take into account the presence of roads) overlaid with road network maps to forecast the spatial distribution of risk of roads becoming blocked by landslides are introduced.

Density of Landslides in Proximity to Roads

By overlaying landslide inventories with buffered road network maps, it is possible to investigate the link between road presence and landslide occurrence by looking at the density of landslides within various buffer zones of the road. For instance, Larsen and Parks (1997) used an inventory of 1,609 landslides that occurred between ~1940 and 1990 in a 201 km^2 forested region of Puerto Rico, where there is 126 km of paved road network across the sloped, forested areas of the study region. Using various width buffers (from

5−400 m), they investigated the density of landslides in each buffer zone, and found that landslide density is five to eight times higher within an 85 m buffer zone either side of the road network than at further distances from the road. Brenning *et al.* (2015) applied this concept to an inventory of 2,185 landslide initiation points within a 300 m buffer zone of highways in Southern Ecuador. They found that landslide incidence was approximately one order of magnitude higher within a 150 m buffer zone of the road network compared to the zone of all greater distances from the road network. Other authors have found increases in landslide incidence within buffer zones close to the road networks (compared buffer zones at greater distances from the road network, or the whole region on average) in either (i) low- to middle-income countries or (ii) 'unimproved' roads (i.e., minor roads or logging roads) (Sidle and Ochiai, 2006; Miller and Burnett, 2007; Das *et al.*, 2012; Muenchow *et al.*, 2012). Authors typically attributed this increase in landslide incidence near roads to poor road design and drainage. Muenchow *et al.* (2012) and Brenning *et al.* (2015) also found that landslides were larger, deeper and more frequently reactivated in the buffer zone around highways, which they attributed to the landscape not recovering between successive slope failures (e.g., through re-vegetation). Several authors (including Brenning *et al.*, 2015) have developed this methodology further to create landslide susceptibility maps where proximity to roads is used as a variable, which is now discussed.

Including Road Networks in Susceptibility Models

Above, examples where authors had investigated the density of landslides in proximity to the road were introduced. These studies assess whether the presence of roads increases landslide occurrence (Sidle and Ochiai, 2006; Miller and Burnett, 2007; Das *et al.*, 2012; Muenchow *et al.*, 2012; Brenning *et al.*, 2015), which was shown to be the case for several examples in low- to middle-income countries or on unimproved roads.

From a database of articles recently collated in a critical literature review of 426 peer-reviewed articles published on the topic of landslide susceptibility (Malamud *et al.*, 2014), we found that 109 of these articles used proximity to roads as a thematic variable in landslide susceptibility map production (highly cited examples include: Chung and Fabbri, 1999; van Westen *et al.*, 2003; Ayalew and Yamagashi, 2005). Of these 109 articles, 85 % were creating susceptibility maps for low- to middle-income countries (where national economic status corresponds to rankings by World Bank, 2014). In a sample of articles where the relationship between landslides and proximity to roads was presented in detail within the paper (n = 6), authors consistently found that landslide density (defined by the

number of landslide pixels in a zone/total number of pixels within a zone) was higher within a 100 m zone either side of the road than at greater distances from the road, and landslide density generally decayed by buffer zone of increasing distance from road (Oh and Pradhan, 2011; Akgun, 2012; Xu *et al.*, 2012; Shahabi *et al.*, 2013; Zare *et al.*, 2013; Remzi and Aydin, 2014). An example of a landslide susceptibility map including distance from roads as a variable in susceptibility map production is shown in **Figure 7.3**.

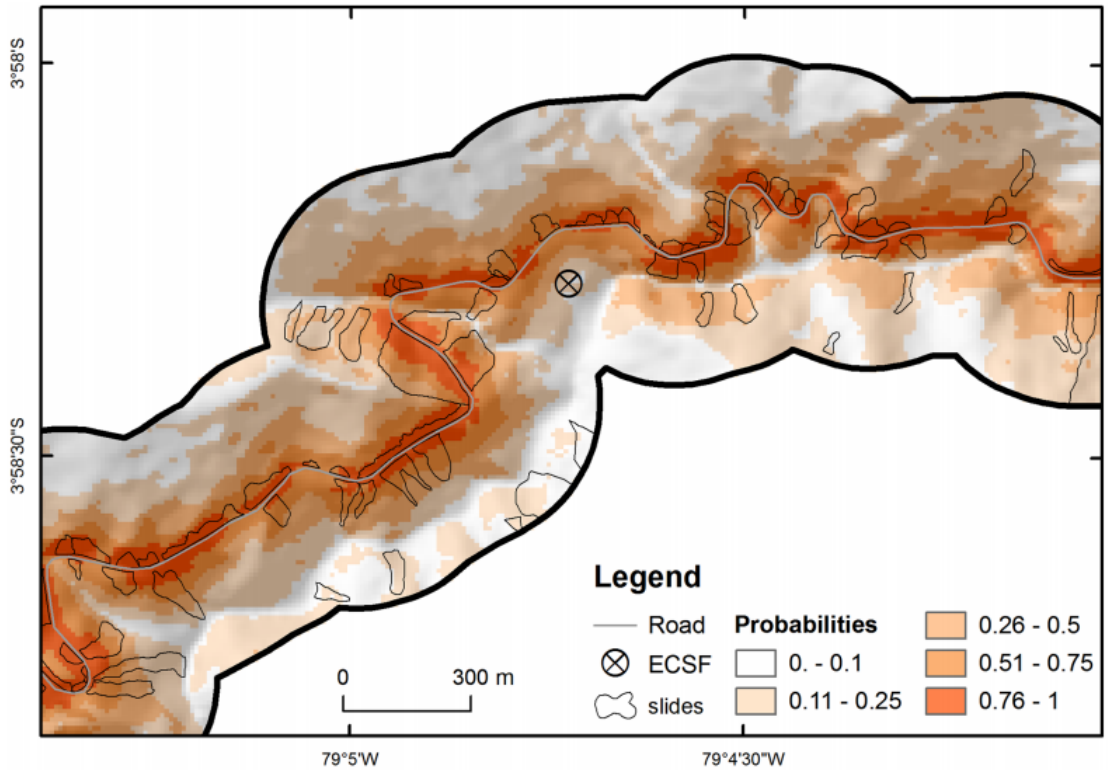


Figure 7.3 *Subset of a landslide susceptibility map produced for part of the Loja Province, Ecuador where distance from roads is included as a factor determining landslide susceptibility. In this example, proximity to the road increases landslide susceptibility. Figure from Brenning et al. (2015). Also overlaid are observed landslide polygons (black outlines), which are typically located along the road corridor.*

The relationship between landslide susceptibility and proximity to roads in high-income countries (i.e., countries where the road network is primarily paved, well-engineered in terms of cuttings and drainage and there may be in-built landslide road protection measures) is less clear. Of the 15 % of 109 articles reviewed above that used proximity to roads as a thematic variable in the creating of susceptibility maps for high-income countries, these articles were generally focused on quite remote areas, where arguably many of the roads may be 'unimproved'. Arguably, the relative absence of articles that include roads in a susceptibility model that are applied to developed areas in high-income countries could suggest that there is not a strong link between presence of roads and increased landslide susceptibility for these locations, although this is far from proven.

Combining Susceptibility Models and Road Networks

The above paragraphs showed that although many articles use proximity to roads as a thematic variable in landslide susceptibility map production, relatively few outline this relationship in detail, and this analysis of landslide-road interactions is somewhat incidental in the production of a susceptibility map. Some authors have more directly used landslide susceptibility maps to investigate landslide road impact vulnerability and risk by overlaying maps of susceptibly and road networks in a GIS. For example, Winter *et al.* (2008) created maps of debris flow susceptibility for Scotland, which were then overlaid with a road map of the Scottish trunk road network to identify stretches of road particularly at risk from debris flows. This country-wide map was then used to prioritise sites to investigate in more detail based on field surveys and socio-economic impact analysis. A similar technique was applied by Quinn (2010) to a region in Eastern Canada.

Remondo *et al.* (2008) expanded on this concept for a 140 km^2 region in Northern Spain to calculate a map of probability of landslide occurrence over the next 50 years. This was performed by combining a landslide susceptibility map with an inventory of landslides from the previous 50 years (assuming no change in temporal probability of landslides, and a single value of landslide area). They calculated a second map of vulnerability based on collecting information on losses caused by landslides over the past fifty years from sources such as the insurance industry, public consultation and transport departments. The map of vulnerability expresses the economic loss caused by a landslide impacting a road segment divided by the economic value of that road segment. Remondo *et al.* (2008) combined these maps of probability of landslide occurrence and vulnerability to calculate the projected risk to the road network caused by landslides on a cell by cell basis, measured in euros per 50 year period. The authors also performed some economic analysis of the indirect losses caused by road blockages in terms of loss of working time and cost of journey detours (assuming that a motorway will be shut for 2 hours, major roads shut for half a day for and local roads shut for one day). Remondo *et al.* (2008) estimated that total risk (direct and indirect losses caused by landslide) per 10 × 10 m pixel vary from €00.00–€60.52 per 50 yr, which equates to an estimated cost of landslide impact across the study area (i.e., all pixels aggregated) amounting to €1,166,639 (in 2008 euros) over the next 50 years. However, this is based on an incomplete inventory of landslides and losses, and it is also not clear whether simultaneous road blockages were taken into account. Zêzere *et al.* (2007) also applied a technique similar to Remondo *et al.* (2008) to the 34.4 km A9 motorway in

Portugal showing a range of cost estimations dependent on how long the motorway was closed for (between 1 and 6 months).

Jaiswal *et al.* (2010a,b) and Jaiswal *et al.* (2011) built upon Remondo *et al.* (2008)'s study for a 22 km^2 area in Southern India by combining a landslide susceptibility map with analysis of landslide frequency-magnitude and return periods calculated from an inventory of 1,084 landslides. Jaiswal *et al.* (2011)'s data inputs and workflow are shown in **Figure 7.4**. Using the inventory of landslides, they calculated the number of landslides expected per kilometre of transportation corridor for different return periods. Frequency-size statistics of inventoried landslides were calculated (as per **Chapter 4** of this thesis) to calculate the probability of a landslide of a given magnitude. Final landslide risk was presented as the number of landslides of a given magnitude within a given return period per kilometre of road. From this, direct risks such as the probability of a vehicle being struck and indirect risks such as additional fuel costs from detours and loss of business approximated. Costs of detours were calculated based on the length of the alternative route, average daily number of travellers, cost of fuel and time of day, and loss of business approximated from participatory interviews with local businesses. Jaiswal *et al.* (2010a) estimated total direct and indirect loss over a 50 year period to be US$779,500 (we assume this is calculated to approximately 2010 relative US$ values).

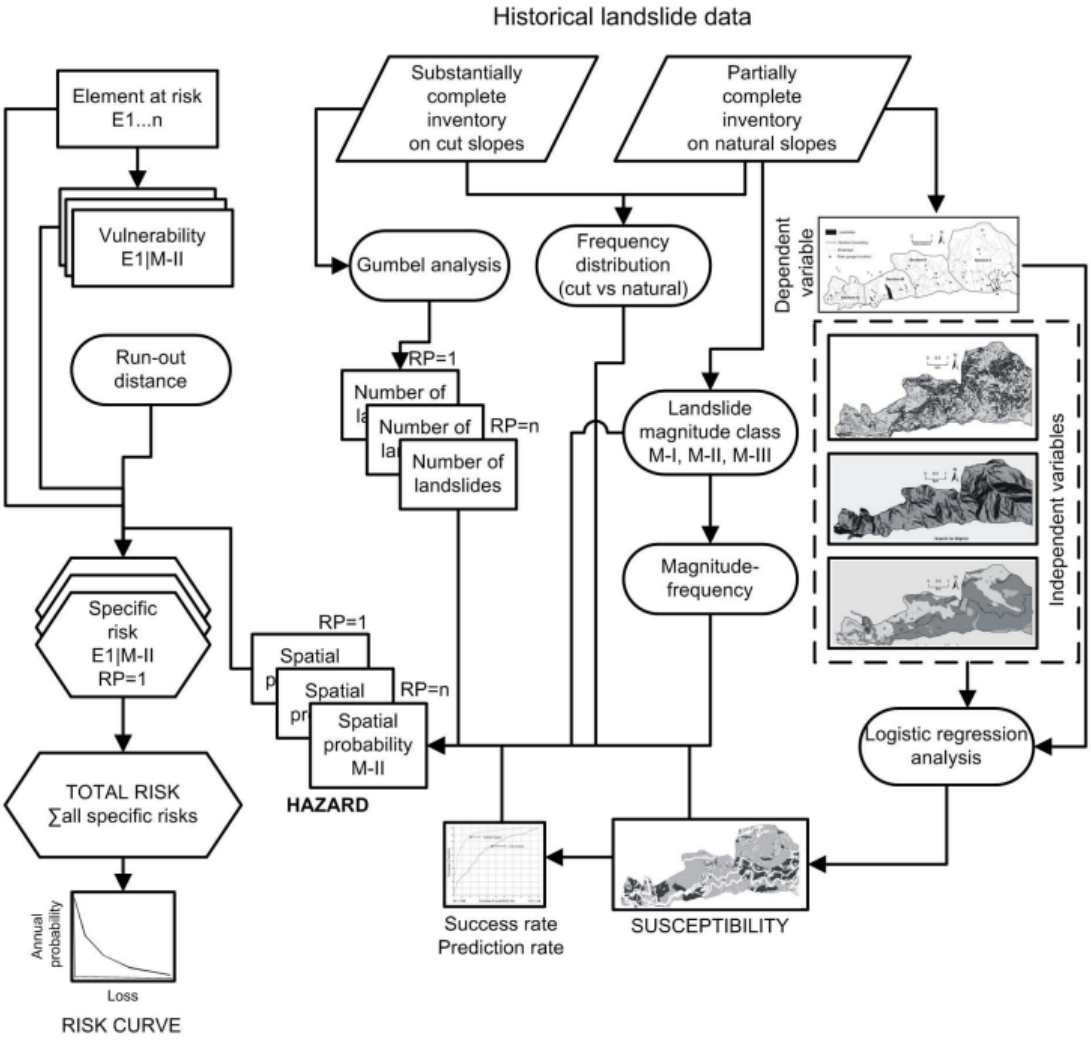


Figure 7.4 *The data inputs and workflow for Jaiswal et al. (2011)'s methodology to estimate landslide risk for a 22 km² area in Southern India. This figure illustrates the large volumes of landslide, environmental and social data required for this calculation of risk. Figure from Jaiswal et al. (2011).*

The methods of creating a map of landslide susceptibility and overlaying with a road network map discussed in this section are similarly applied in the rockfall analysis literature. This literature is not examined in depth here, as LRIM focuses on modelling low mobility landslides in this thesis (i.e., all landslides other than rockfalls and debris flows). An illustrative example is Guzzetti *et al.* (2004b) who created a rockfall hazard map by repeatedly running a rockfall simulation model to identify zones where modelled rockfall trajectories most commonly occurred. This was overlaid with the road network map to identify high risk stretches of road and combined with analysis of average daily traffic flow to calculate how much time vehicles spend in these high risk zones, and the consequent risk of vehicles being struck by rockfalls.

Although the studies described here are each comprehensive in calculation of total economic risk caused by landslides blocking the transportation network, these studies

highlight the very labour intensive methodology required to (i) collect/create data on landslide occurrence and susceptibility and (ii) understand and measure travel choices, behaviour and economic losses and collect data from various public and private bodies. All studies acknowledged that the economic estimations they produced were likely to be conservative due to (i) incomplete landslide inventories used in analysis and (ii) difficulty in obtaining consistent loss estimates for all landslides occurring over a number of decades. As a consequence, this methodology (identifying high risk portions of road using a susceptibility map and exploring scenarios of road disruption) is typically applied to small regions or individual stretches of road where the number of variables can be reasonably constrained (e.g., the number of alternative travel routes).

7.2.2.5 Traffic Impact Modelling

Section 7.2.2.4 gave examples of using landslide susceptibility models overlaid with road network maps to calculate the long term economic risk caused by both direct road damage and indirect impacts such as detours and loss of business. More recently, a small number of studies have borrowed from the transportation literature (discussed later in **Section 7.2.3.2**) to assess the impact of landslide-road interactions upon traffic. For instance, Voumard *et al.* (2014) developed a dynamic traffic simulator applied to three small stretches of road (of the order of 1 km length each) in the Swiss Alps susceptible to different types of landslide. Rather than using daily average values of traffic flow through each stretch in the calculation of risk (as is normally the case), Voumard *et al.* (2014) modelled each individual car's movement through the stretch of road, where speed varied as a function of road conditions (e.g., slowing down for hairpin bends, queueing if a landslide had already blocked a road). Through this finer scale modelling of traffic flow, Voumard *et al.* (2014) demonstrated that risk of death by a vehicle being struck by a landslide was up to 150 % higher for certain stretches of road where there were a combination of difficult driving conditions and high probability of landslide incidence.

At the opposite end of the scale, Meyer *et al.* (2015) showed a methodology to calculate annual additional traffic load across all local and major roads in Southern Norway caused when a debris flow blocks and closes a road, resulting in detours. They first computed the shortest route between ten major cities across the region, and then calculated the annual likelihood of road closure for all roads by combining a susceptibility map with analysis of the debris-flow trigger frequency (illustrated in **Figure 7.5A**). For each route connecting two of the ten major cities, they then calculated the next shortest path between those cities,

and effectively redistributed the average daily traffic flow to this alternative path to calculate the additional demand upon these roads (**Figure 7.5B**) and the additional fuel costs of these detours. They found that some of the roads most susceptible to blockage (in the mountainous central region) were in regions of very low road network density, resulting in detours of up to 200 km, resulting in tens to millions of additional vehicles per kilometre of alternative route.

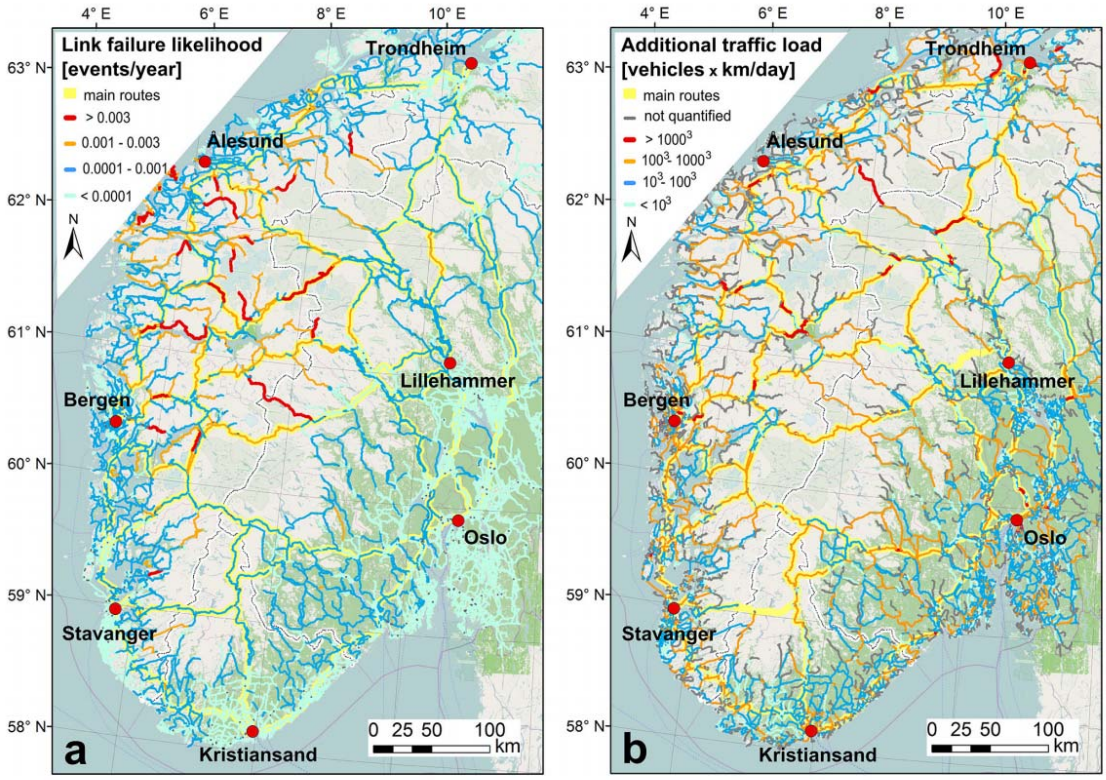


Figure 7.5 *Meyer et al. (2015)'s calculation of traffic impact of debris flows on routes connecting 10 major cities of southern Norway.* ***(A)*** *Calculation of annual probability of road network blockage caused by debris flows for each road network link.* ***(B)*** *Additional traffic load on alternative routes if a link becomes blocked. Figure from Meyer et al. (2015).*

Although the studies described in this **Section 7.2.2.5** are subject to uncertainties and methodological flaws, they illustrated new and useful ways of measuring risk of landslide-road impact and/or revealed that current estimates of risk may be underestimated. Arguably these types of (i) very fine scale analysis of individual cars and (ii) large scale analysis of country wide traffic, would have been considerably more difficult to perform a decade ago with less computing power and less user-oriented network analysis tools and algorithms available (Reggiani *et al.* 2015). However, as computing power increases; new regions of the world open up to international trade via road building (Fan and Chan Kang, 2005; Garver, 2006) and more established regions become more risk-averse (Cha and Ellingwood, 2012), we believe there will be increasing synergy between the hazards and

network analysis research communities to produce robust estimates of landslide-road impact risk. **Section 7.2.3** discusses some of the concepts from the transportation modelling literature that could potentially be applied to the problem of Landslide-Road Impact Modelling.

7.2.2.6 Summary of Current Techniques used to Understand Landslide-Road Impact

Section 7.2.2 showed that currently, our main methods for understanding and forecasting the impact of landslides on road networks are:

(i) *Forensic cost studies and economic models* where the cost of previous events is extrapolated or scaled to forecast the regional impact of future events.

(ii) *Overlaying landslide inventories over road network maps* to identify the locations and sizes of landslide-road blockages.

(iii) *Use and creation of landslide susceptibility maps* to understand whether road building increases likelihood of landslides occurring and also to spatially identify portions of road that intersect high landslide susceptibility zones.

(iv) *Traffic impact modelling* to understand the delays and detours caused by landslides on the road network in a small number of recent papers.

Some key findings from the literature relevant to the development of the Landslide-Road Impact Model (LRIM) here:

- *The costs of repairing roads* damaged by landslides and the indirect cost of denial of access to locations and facilities are large and typically underestimated (e.g., Wills *et al.*, 2014).
- *Forensically reconstructing the economic cost* of previous landslide impacts upon a road network is a labour-intensive process in terms of data collection from various bodies, and is perhaps prohibitively 'expensive' to perform for many regions of the world where landslide road impact is a problem (Klose *et al.* 2014).
- *Spatially overlaying landslide inventories and road network maps* allows us to derive simple and useful indices to measure multiple road network blockages and compare between different events (e.g., total number of road blockages, the average length of road between road blockages) (Guzzetti *et al.,* 2003).
- *Lack of landslide inventories for a region* may make it difficult to understand the spatial interaction between landslides and roads. Records of landslide-road impact may be

kept by various road maintenance agencies and local governments and not recorded in a universal, systematic way (Wills *et al.*, 2014).

- *Road building has been shown to increase the incidence of landslide occurrence relative to surrounding areas* where there are no roads. This has been shown mainly for 'unimproved roads', particularly in low- to middle-income countries, and is attributable to poor drainage and slope modification (Brenning *et al.*, 2015). The link between landslide occurrence and road presence for 'improved roads' in high-income countries is less clear.
- *Landslide susceptibility maps* are commonly used to forecast the spatial occurrence of landslides. Around 25 % of papers on the topic of landslide susceptibility map creation reviewed by Malamud *et al.* (2014) use proximity to roads as a variable that increases landslide susceptibility. Around 75 % of papers did not investigate the link between road presence and landslide occurrence.
- *Landslides are typically considered as 'individual' events rather than populations*. Most studies tend to either implicitly or explicitly focus calculating the risk and impact of *one* landslide blocking a road (sometimes with an analysis of temporal probability) rather than *multiple* landslides simultaneously blocking roads. This is likely to mean that risk and impact are underestimated for triggered landslide events where there is the potential for multiple road blockages to occur.
- *It is possible to borrow techniques from the transportation literature to understand the impact of landslides on traffic flow*. In the last two years, a small number of papers have emerged that borrow techniques from the transportation literature to more actively model the traffic implications of landslide-road network impact (e.g., Voumard *et al.*, 2014). To some extent, this can be attributable to increases in personal computing power and increasingly user-oriented network analysis software.

Other bodies of literature assessing landslides on the road network not examined in detail here include:

(i) *Modelling of rockfalls*. As rockfalls are not included in the LRIM, this literature was not examined in full detail although key methodologies were discussed in **Section 7.2.2.4** (e.g., Guzzetti *et al.*, 2004b).

(ii) *Engineering literature* discussing road building and slope stabilisation techniques to reduce the risk of landslides occurring on roads (e.g., Turner and Schuster, 1996 and Hearn *et al.,* 2011). As the LRIM model approach is more statistical

than geotechnical, we have not investigated this literature in detail. However, some of this literature is referred to in **Section 7.3.2.6** where we discuss regional approaches to road building which may require more local parameterisation of LRIM.

This **Section 7.2.2** has highlighted the labour-intensive nature of data collection to understand landslide-road impact, which typically involves the creation of detailed landslide inventories, databases and susceptibility maps and engagement with various public and private bodies. From personal experience working on the National Landslide Database of Great Britain (**Chapter 3**), in a country with relatively low landslide occurrence and relatively high resources available for risk assessment, the effort involved to liaise with local councils, highways agencies, and private landowners in any systematic way to record landslide-road impact is not feasible. Thus a second aim of the LRIM model methodology presented here is to minimise the local data collection required to operate the model.

Section 7.2.2.6 showed that more recently, a handful of studies have emerged using network analysis techniques looking at the network wide impacts of individual road-blocking landslides, and it is an aim of the LRIM methodology presented here to build upon this work to examine the probability of simultaneous failures across a regional road network during a triggered landslide event and implications in terms of accessibility across the network. The following **Section 7.2.3** briefly discusses some of the broader network analysis literature that we hope to borrow techniques from for the LRIM model.

7.2.3 Natural Hazard Impact Modelling from Other Disciplines

Section 7.2.2 demonstrated that much of our understanding of landslide-road network impact is focused on regional studies, investigating the impact of single (or a small number of) landslides and often requiring large amounts of locally specific data collection. Across other hazard research disciplines, there are several approaches which we believe could be applied to the development of the Landslide-Road Impact Model (LRIM) to allow (i) forecasting of impact with minimal data requirements and (ii) the investigation of multiple simultaneous road blockages in a triggered landslide event. **Section 7.2.3.1** discusses examples and methods in the literature aimed at generally applicable road network impact analysis and **Section 7.2.3.2** discusses examples of simultaneous road network impacts.

7.2.3.1 Generally Applicable and Low-Data Input Hazard-Road Impact Models

The FEMA HAZUS Model: Flexible and Generally Applicable

The United States Federal Emergency Management Agency (FEMA) HAZUS spatial model of risk and loss (FEMA, 2003) is perhaps the best example of a standardised approach and generally applicable model to forecast the impact of various hazards upon buildings and infrastructure. The HAZUS model simulates different scenarios of flooding, earthquakes and hurricanes and can be applied to any location in the United States of America (Schneider and Schauer, 2006). It has also been adapted and applied to locations outside the USA (e.g., Hansen and Bausch, 2007). The HAZUS model can be run in different modes, depending on user expertise and amount of data available: from an 'out of the box' set up, using standard in-built databases to a more advanced, adaptable model with a broader range of outputs if more locally specific data is available, providing flexibility for application in different locations (Scawthorn *et al.*, 2006). HAZUS is a combination of deterministic modelling of the hazard (e.g., for flooding, potential water depth is calculated from a DEM; for earthquakes, ground motion is calculated from historical source parameters and attenuation relationships) overlaid with maps of elements at risk (e.g., buildings and roads by type). The amount of damage for each element is based on fragility curves linking the magnitude of a hazard at that location to the amount of damage likely to be caused. These fragility curves are constructed from a combination of empirical relationships between magnitude of previous events and observed damage and physical modelling (Kircher, 2006). Damage caused to each individual element can then be summed to calculate the overall impact of an earthquake, floor or hurricane event across a region. For the transportation network, only the direct costs of repair are currently calculated within HAZUS.

The accuracy of forecasts from the HAZUS model is strongly linked to the level of data provided (Khater *et al.*, 2003), and forecasts based on the 'off the shelf' model are known to be have a higher degree of uncertainty than forecasts when locally specific data is available (Neighbors *et al.*, 2012), but do provide a 'rough' first estimate in a standardised way from which decision makers choose to perform more in-depth analysis. HAZUS has been developed by a large body of scientists over at least two decades, and has the benefit of large, countrywide databases of elements at risk and past damages, so it is beyond the scope of this thesis to develop a comparable tool for landslide road impact estimation here, particularly in terms of the creation of robust fragility curves. However, the principles of (i)

using relatively simple, 'universal' behaviour of a hazard to create a widely applicable, standardised risk assessment and (ii) flexibility of the tool to allow the incorporation of more specific local data and expertise if available are adopted here.

Monte Carlo Modelling Techniques

The previous paragraphs discussed the idea that the more 'generic' the data and approach, the more uncertain the model output will be. One method commonly used in the catastrophe insurance loss modelling literature to quantify and somewhat deal with this uncertainty is by performing Monte Carlo simulations (Grossi and Kunreuther, 2005). In the setting of this thesis, Monte Carlo type simulations involve randomly sampling values from established probability density functions (pdfs) and repeating this process hundreds to thousands of times (Huang *et al.*, 2001). The pdfs correspond to a damage-causing characteristic of a hazard (e.g., wind speed for hurricanes, ground acceleration for earthquakes, water depth for floods) or set of parameter values to model the hazard and may be generated based on previous observations (as discussed with regard to landslides in **Chapter 4**). Thus, by repeatedly sampling from a pdf, thousands of different hazard scenarios are explored, from 'extreme' events in the tails of the pdf to more common events around the centre of the pdf (although different types of 'common' event may result in very different impacts, e.g., if two hurricanes of the same wind speed make landfall in a city versus a sparsely populated area). Examples include Musson (2000) for earthquakes, Pinelli *et al.* (2004) for hurricanes and Merz *et al.* (2004) for flooding. Monte Carlo approaches may also be used to investigate variability in the impact for a specific hazard event scenario. For example, Nojima and Sugito (2000) created one scenario of regional ground motion from an earthquake, which they overlaid with a road network map where each link of the road network had a fragility curve linking shaking intensity to probability of failure. The authors then adopted a Monte Carlo approach to vary which links of the road network actually broke, creating thousands of different road network interruption patterns. They then used a network analysis technique (which is discussed in the following **Section 7.2.3.2**) to measure the regional impact on the transportation network for each Monte Carlo simulation, and analysed overall results to identify commonly occurring patterns of network disruption.

7.2.3.2 Network Analysis Techniques

To efficiently measure the impact of simultaneous network impacts from a hazard and the resulting change in accessibility across a region, many studies use network analysis indices

and techniques from the graph theory literature. Graph theory captures the topology of the network structure (Andrews and Estabrook, 1971). In the graph theory approach, roads are represented as straight lines (referred to as edges, arcs or links) which are assigned a value depending on the nature of the study (e.g., distance, traffic flow, speed). These vertices are connected by points (referred to as nodes) which can represent features such as origins/termini, intersections or settlements on the network. The resulting 'graph' is a visual representation of the linkages between points although the spatial layout does not represent the geographic locations of the features.

The visual graph is somewhat arbitrary as the nodes'/vertices' information is stored and processed in a range of matrices which allow rapid processing of mathematical indices for measurement of the road network (West, 2001). **Figure 7.6** and **Table 7.2** demonstrate an example of the simplification of the railway network of Sardinia to a network diagram and adjacency matrix (Kansky, 1963). From the abstraction of the road network using graph theory, it is possible to derive a range of indices to measure the road network and changes to it as a whole.

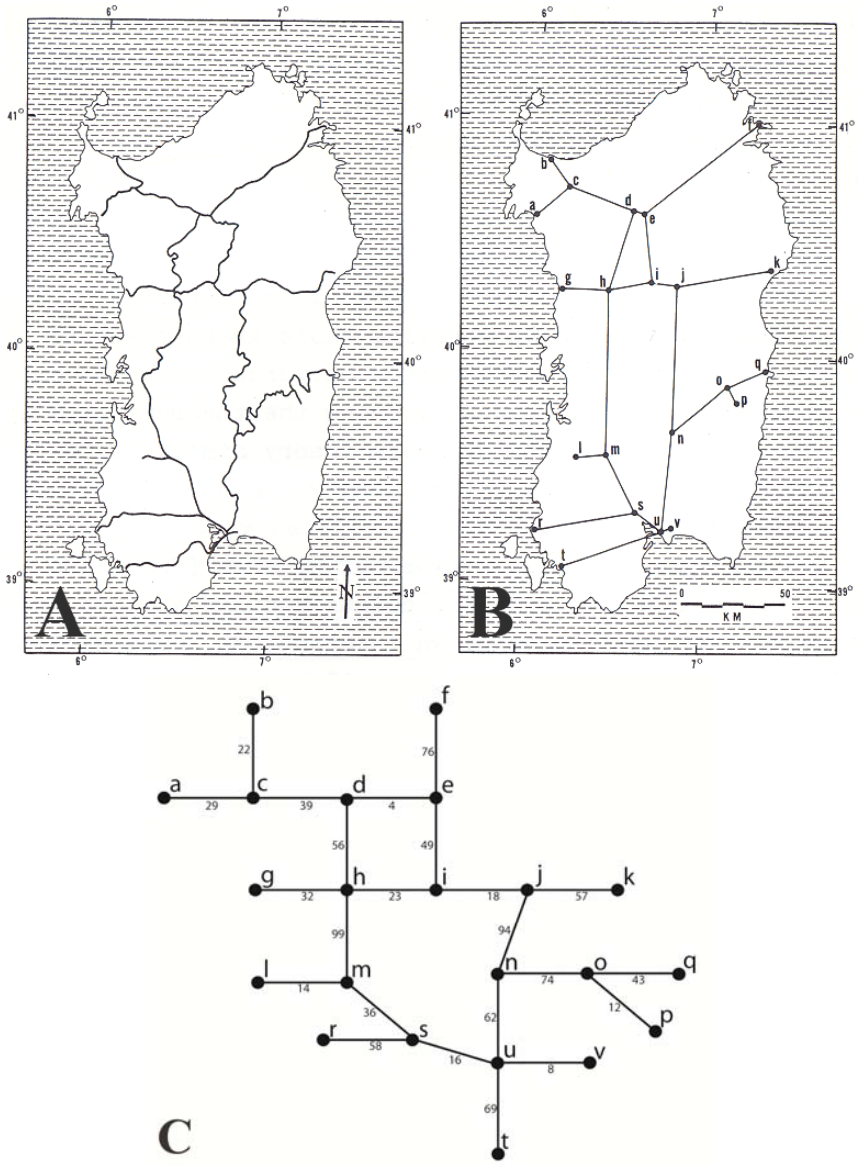


Figure 7.6 *Representation of the railway network of Sardinia using network theory where nodes are lettered and edges represent distances between nodes. (**A**) Map of the railway network of Sardinia in geometric form. (**B**) Conversion of map A to edges (railways) and nodes (stations). (**C**) Simplification of map B where all edges are visually the same length but are assigned different values depending on the distance between nodes. Figures from Kansky (1963) with Figure C adapted.*

Table 7.2 *Example of a simple adjacency matrix for stations (nodes) of the railway network of Sardinia, demonstrated in* ***Figure 7.6****. Here, each station (node) is identified with a letter (a to v). If two nodes are connected by a single edge, the distance between those two nodes is recorded in the adjacency matrix. Where there is no direct link between nodes, an arbitrary value of '0' is given. This matrix allows rapid calculation of travel distance between any nodes on the graph by addition of distances between individual nodes. Although application of graph theory does result in a visual representation of the network, much of the analysis is carried out in the accompanying matrices. Matrices will differ in terms of structure and complexity depending on their purpose. Table constructed with distance values from Kansky (1963).*

	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v
a	0	0	29	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
b	0	0	22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
c	29	22	0	39	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
d	0	0	4	0	4	0	0	56	0	0	0	0	0	0	0	0	0	0	0	0	0	0
e	0	0	0	4	0	76	0	0	49	0	0	0	0	0	0	0	0	0	0	0	0	0
f	0	0	0	0	76	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
g	0	0	0	0	0	0	0	32	0	0	0	0	0	0	0	0	0	0	0	0	0	0
h	0	0	0	56	0	0	32	0	23	0	0	0	99	0	0	0	0	0	0	0	0	0
i	0	0	0	0	49	0	0	23	0	18	0	0	0	0	0	0	0	0	0	0	0	0
j	0	0	0	0	0	0	0	0	18	0	57	0	0	94	0	0	0	0	0	0	0	0
k	0	0	0	0	0	0	0	0	0	57	0	0	0	0	0	0	0	0	0	0	0	0
l	0	0	0	0	0	0	0	0	0	0	0	0	14	0	0	0	0	0	0	0	0	0
m	0	0	0	0	0	0	0	99	0	0	0	14	0	0	0	0	0	0	36	0	0	0
n	0	0	0	0	0	0	0	0	0	94	0	0	0	0	74	0	0	0	0	0	62	0
o	0	0	0	0	0	0	0	0	0	0	0	0	74	0	0	12	43	0	0	0	0	0
p	0	0	0	0	0	0	0	0	0	0	0	0	0	0	12	0	0	0	0	0	0	0
q	0	0	0	0	0	0	0	0	0	0	0	0	0	0	43	0	0	0	0	0	0	0
r	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	58	0	0	0
s	0	0	0	0	0	0	0	0	0	0	0	0	36	0	0	0	0	58	0	0	16	0
t	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	69	0
u	0	0	0	0	0	0	0	0	0	0	0	0	0	62	0	0	0	0	16	69	0	8
v	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	8	0

Broadly the type of graph theory analysis of hazards on road networks illustrated in **Figure 7.6** and **Table 7.2** is aimed at identifying network elements most likely to fail, modelling the resulting changes to traffic patterns due to network disruption, or modelling economic losses associated with network disruption. Patterns of network disruption are most often generated through developing (or using existing) fragility curves to calculate the probability of network element failure for different magnitude events and using these in Monte-Carlo simulations to derive the 'most likely' elements to fail for a given event scenario. Traffic indices such as increased journey time or distance are then calculated from the shortest route through the network between pre-defined origin and destination pairs. For example, Chang and Nojima (2001) simulated the impact of the 1995 Kobe Earthquake on the road and rail systems of the Kobe, Japan region using indicators such as length of the network open to show recovery of the road network over a 24 month period (illustrated in **Figure 7.7**). In a second example, Fiedrich *et al.* (2000) developed a model to efficiently allocate resources (such as rescue equipment) across a network following an earthquake disruption. In a third example, Jenelius and Mattsoon (2012) created a model for 'generic' hazard network disruptions lasting hours to days to identify particularly vulnerable road network links in terms of traffic demand.

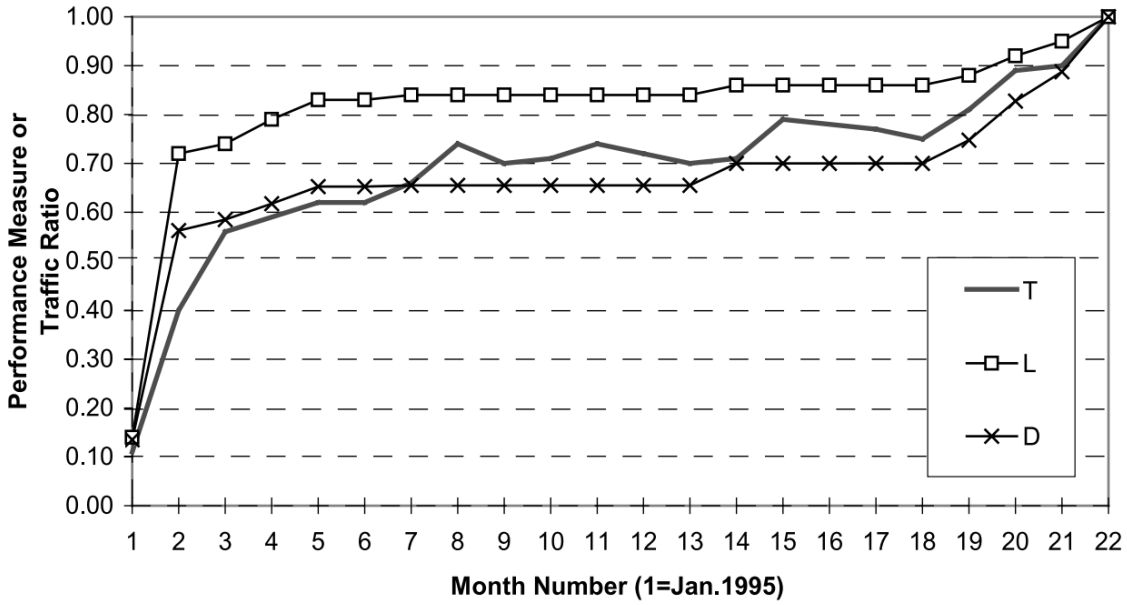


Figure 7.7 *Network indices (T, L and D) used to measure the recovery of the Kobe road network over a 24 month period following the Kobe earthquake. T is a measure of total daily traffic flow across the road network, based on average traffic flow per road segment, if a road segment is open (i.e., not damaged by the earthquake). L is a measure of the length of the road network open (i.e., the sum of link lengths) and D is a measure based on how easy it is to move about the network (derived from the average travel distance between pairs of nodes, which will be higher if detours are required). The Kobe earthquake occurred in Month 1, resulting in damage to the road network (as shown by low values of T, L and D). Over the 24 month period (ending December 1996), the functionality of the regional road network gradually recovers from earthquake damage, shown by increases in T, L and D parameters. Figure from Chang and Nojima (2001).*

The graph theory approach also aids in the a-priori analysis of network vulnerability and/or resilience (e.g., prioritising where new roads should be built to avoid locations becoming isolated, or identifying which road links would have the greatest impact on travel in a region if they became blocked). For example, Sohn (2006) developed a model to simulate the impact of flooding on regional highways and identify stretches of road most suitable for retrofitting and protection measures. Network topologies resilient to failure are generally those with high values of redundancy (several alternative routes between points) and those with high values of clustering (sub-regions where all nodes are connected to one another) (Cats and Jenelius, 2015). However, road networks are somewhat different to other infrastructures (such as water, electricity and communications) as they are considerably more spatially limited, e.g., there are rarely more than four or five roads joining together at a road junction (Jenelius *et al.*, 2006). Consequently optimising road network topologies has more limitations than other networks. Moreover, there are few instances where an entire road network is designed and installed 'from new', thus the focus is more towards planning for failure across existing network elements than designing 'optimal' road networks.

In the research presented here, we attempt to borrow from these graph theory techniques used in other fields of hazard modelling to create a regional road network impact model for triggered landslide events.

7.2.4 Summary of Existing Research and Future Research Directions for Understanding Landslide Road Impact

The research of other authors identified in **Section 7.2.2** generally did not consider the network wide impacts of road blockages, but instead focused on specific stretches of road to establish:

- *The annual probability of landslides occurring* on that road.
- *The return period* of different magnitude events.
- *The economic implications* of road blockages.

Methods generally involved:

- *Overlaying maps of landslide inventories with contributing factors* (e.g., slope angle, geology, drainage etc.) to produce risk maps (e.g., Winter *et al.*, 2005; 2008).
- *Overlaying landslide inventory maps with road networks to establish the frequency of road blockages* (e.g., Reichenbach *et al.,* 2003).
- *Modelling of economic losses caused by repairs and delays* (e.g., Hearn *et al.,* 2007).

The lack of literature investigating the network wide impacts of a landslide event is somewhat surprising but possibly may be explained by the premise that conceptualising landslides as 'populations' is a relatively new research area (illustrated by the fact that the majority of the references in **Chapter 4** are from the last 10–15 years), and that many regions susceptible to large triggered landslide events are in low- to middle-income countries, where research and civil protection priorities may be geared to more fundamental tasks, like the creation of landslide susceptibility maps.

Section 7.2.3 reviewed some approaches used in other disciplines (particularly the earthquake and transportation literature) which may possibly be adapted to the modelling of triggered landslide events in LRIM. These were:

(i) The use of *relatively simple 'universal' behaviour of hazards* to create generally applicable models.

(ii) *Flexible models* that may allow the incorporation of more data and local expertise if available.

(iii) *Monte Carlo simulations* to quantify and deal with uncertainty.

(iv) *Network analysis techniques* to quantify the impact of a hazard across an entire regional road network.

Here we aim to distil the methods listed in **Section 7.2.2** and **7.2.3** to develop a model that is 'generally' applicable (i.e., not location specific), requires minimal data inputs and looks at the network-wide impacts of landslides occurring on the road network. The following **Section 7.3** gives an overview of the Landslide-Road Impact Model (LRIM) modelling approach and then outline the inputs and methods used.

7.3 The Landslide-Road Impact Model (LRIM)

7.3.1 Landslide-Road Impact Model (LRIM) Overview

Here we consider a Landslide-Road Impact Model (LRIM) for triggered landslide events, where the impact of multiple landslides on the road network, can at times be greater than adding up the impact of individual landslides on specific roads (e.g., where two landslides result in a much larger part of the network being cut off, compared to just one or the other landslide's individual impact). This **Section 7.3** will be organized as follows. We first (this **Section 7.3.1**) give a broad overview of the LRIM. This is followed by discussing the six inputs to the model (**Section 7.3.2**), the pre-loop phases to the model (**Section 7.3.3**), how the model chooses the initial point locations, areas and shapes of the synthetic landslides (**Section 7.3.4**), final placement (**Section 7.3.5**) and orientation (**Section 7.3.6**) of the synthetic landslides, how the model deals with landslides crossing slopes downhill to uphill (**Section 7.3.7**), and finally, overlaying the final synthetic triggered event landslide inventory with roads (**Section 7.3.8**).

In our model methodology, we create synthetic triggered event landslide inventory maps for a given magnitude triggered landslide event chosen by the user (i.e., the number of landslides resulting from the trigger), based on general statistical behaviour of triggered landslide events, combined with local characteristics of susceptibility and terrain. The general statistical behaviour (discussed in **Chapter 4**) refers to the frequency-area distribution for low mobility landslides in many triggered events, which appear to be relatively similar regardless of the triggering mechanism or local characteristics. We also use the primary research presented in **Chapter 5** to realistically model the shape of landslides

as ellipses, where length to width ratio of the ellipse varies with landslide area. Statistical distributions have been developed to describe this behaviour of landslide area and shape, which can be randomly sampled to create synthetic triggered landslide event inventories that have similar characteristics to the real world. The synthetic landslide inventory maps created in our model are then overlaid with the road network to identify where road blockages occur, and where landslides are within 50 and 100 m of the road.

Figure 7.8 presents a broad overview of our Landslide-Road Impact Model (LRIM), including a grouped overview of inputs, methods and outputs.

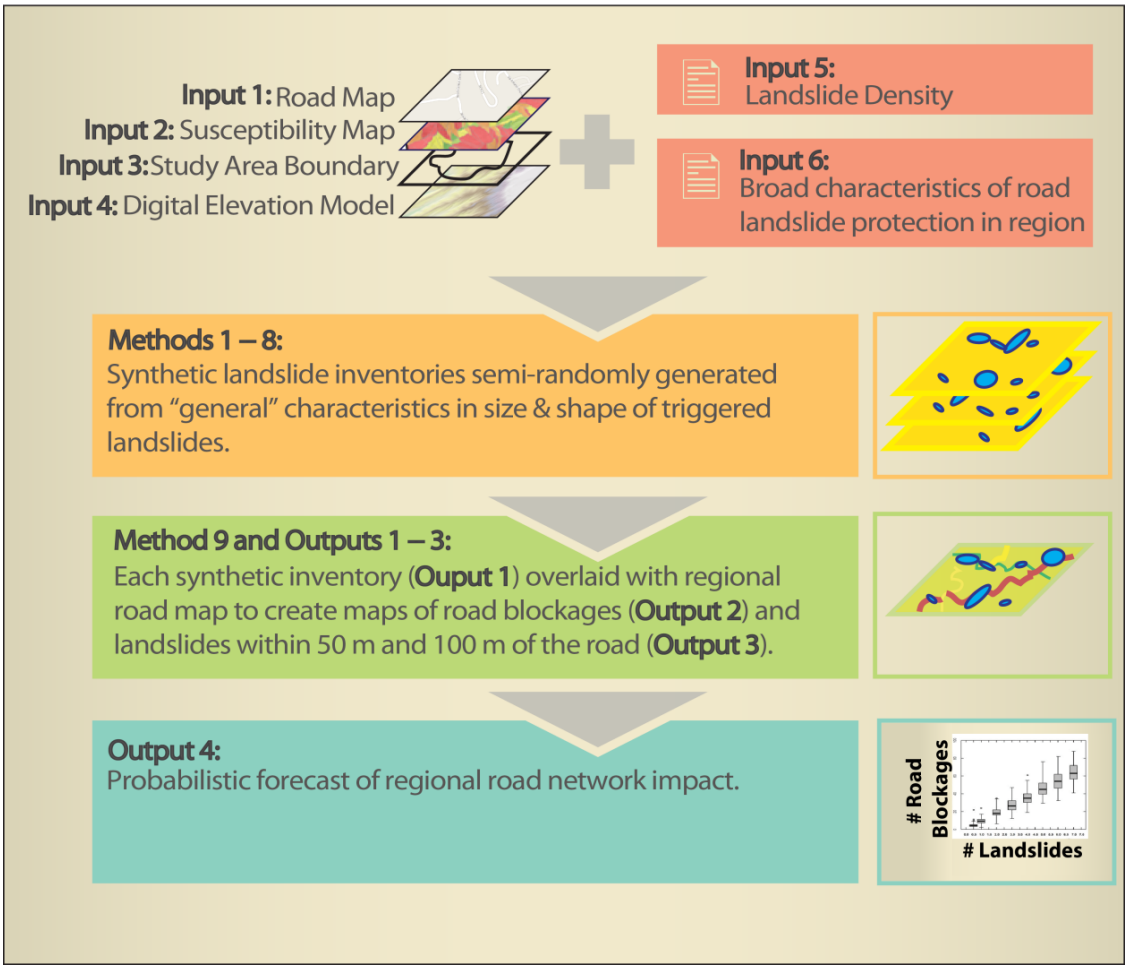


Figure 7.8 *Overview of the Landslide-Road Impact Model (LRIM) process showing user inputs, modelling methods and data outputs.* ***Methods 3*** *to* ***8*** *and* ***Outputs 1*** *to* ***3*** *are repeated multiple iterations (we choose 100) such that multiple synthetic landslide map are produced, for which the probabilistic forecast (****Output 4****) is then determined. For more detail see* ***Figure 7.2****.*

The LRIM, as shown in **Figure 7.8**, is a semi-random process where for a given set of input parameters, multiple synthetic landslide maps are produced using a Monte Carlo simulation (we choose 100 iterations). In each iteration of the LRIM there will be some variability in the synthetic inventory map produced in terms of the sizes, shapes and

locations of individual landslides, and resulting road blockages and landslides within 50 and 100 m of the roads. This Monte Carlo approach allows us to explore the probabilistic range of impacts upon the regional road network, in terms of number, size, location and impact of road blockages occurring after a triggered landslide event. Users may adjust the number of iterations to run the model for. One hundred simulations was chosen here as an easy to manage amount of output data whilst also producing results that are statistically robust (Winston, 1996).

Figure 7.9 presents a more detailed view of the Landslide-Road Impact Model (LRIM), including specifics on each method and output.

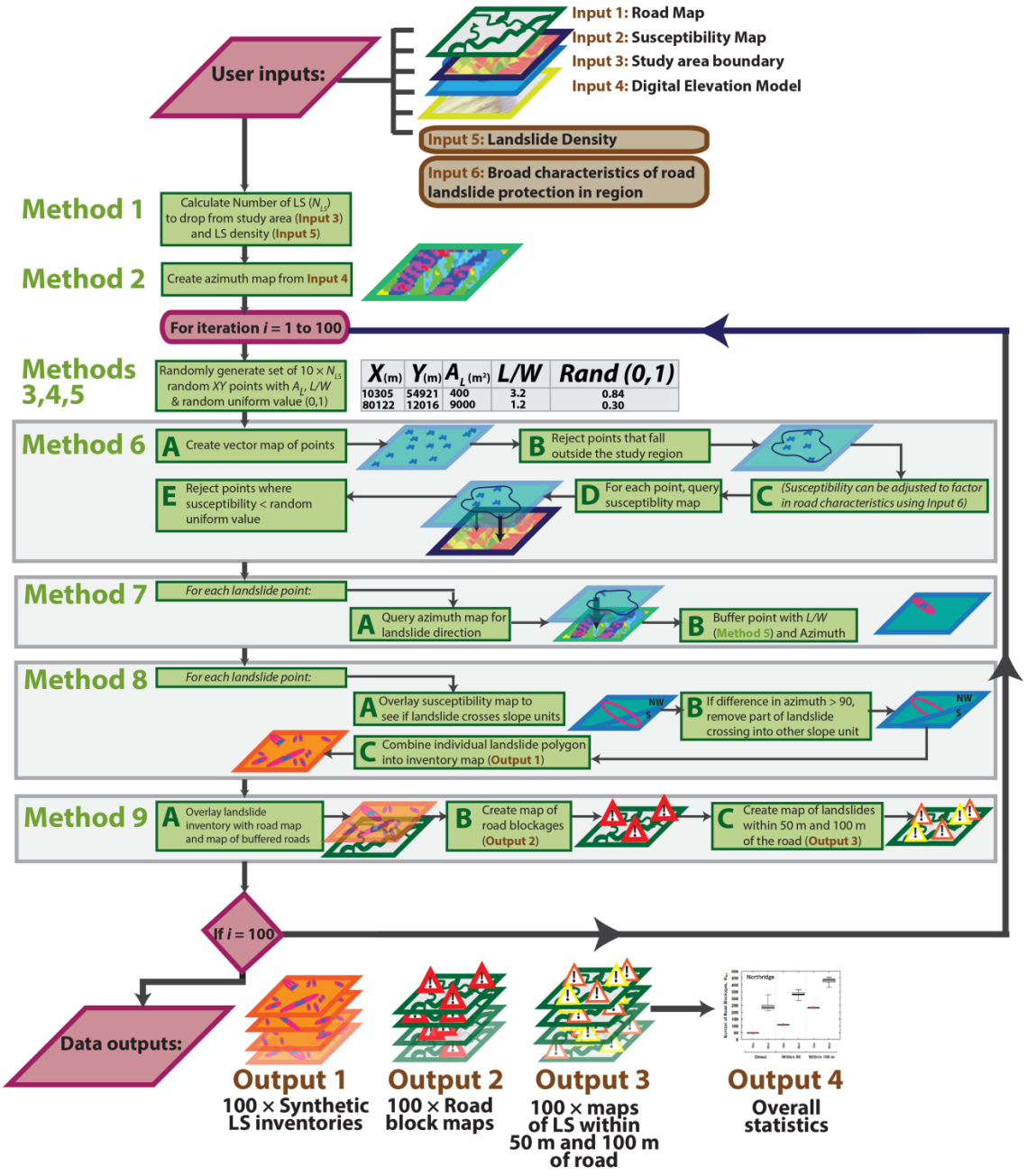


Figure 7.9 *Detailed flow diagram of the Landslide-Road Impact Model (LRIM) process, showing all user inputs, methods used and data outputs from the model.*

The Landslide-Road Impact Model (LRIM) is written in Python programming language and executed in GRASS GIS 7. GRASS GIS is an open source spatial analysis software with many existing tools and add-ons. This allows potential users to adapt this spatial model for their specific location with relatively minimal financial outlay for software. The speed of the simulation will vary depending on the size and detail of the study region, the magnitude of the triggering event and the power of the computer being used, with times to complete 100 iterations of the model varying from a few hours to days. In our own simulations, the model was run in a Linux Ubuntu 64 bit virtual machine with 10 GB of

RAM and 50 GB hard drive. Although the model will run on a standard personal computer, a relatively high powered machine (at least these configurations stated here) is recommended due to the large volumes of data created in the Monte Carlo modelling process.

Because of the generally applicable statistical patterns of landslides used in the model and low cost approach theoretically, the LRIM may be applied to any region of the world susceptible to medium to large triggered landslide events where the majority of landslides are low mobility (i.e., all types of landslide outlined by EPOCH (Flageollet, 1996) excluding rock falls and debris flows). We have applied the LRIM to sites in three sites (see **Section 7.4** for detailed discussion of each site): Collazzone, Central Italy; Oat Mountain, Northridge, California, USA; Su-Hua, Northeast Taiwan.

Sections 7.3.2 to **7.3.8** now discuss in turn the key processes in terms of user inputs and modelling methods, including for each section illustrative examples from the Collazzone region (Umbria, Central Italy).

7.3.2 LRIM Input Data (Inputs 1 to 6)

The first stage of the LRIM process requires the user to point GRASS GIS to the relevant input data files, **Inputs 1** to **4** (**Figure 7.9)**:

Input 1: Road Map

Input 2: Landslide Susceptibility Map

Input 3: Study Area Boundary

Input 4: Digital Elevation Model

The user must also set input parameters, **Inputs 5** and **6** (**Figure 7.9)**:

Input 5: Landslide Density

Input 6: Broad Characteristics of Road Landslide Protection in the Region

User definition of input files is done in a simple graphical user interface, shown in **Figure 7.10**.

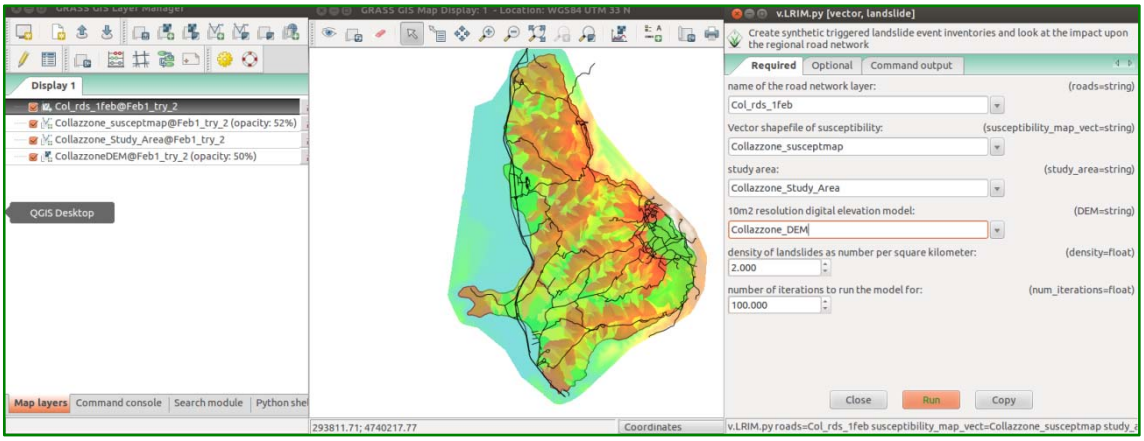


Figure 7.10 *Landslide-Road Impact Model (LRIM) Grass-GIS Graphical User Interface for* ***Inputs 1*** *to* ***4****.*

Figure 7.11 shows an overview of LRIM for **Inputs 1** to **4** for the Collazzone region (Umbria, Central Italy) including a road map (**Input 1**, **Figure 7.11A**), landslide susceptibility map (**Input 2**, **Figure 7.11B**), study area boundary (**Input 3**, **Figure 7.11A** to **C**), and a digital elevation model (**Input 4**, **Figure 7.11C**).

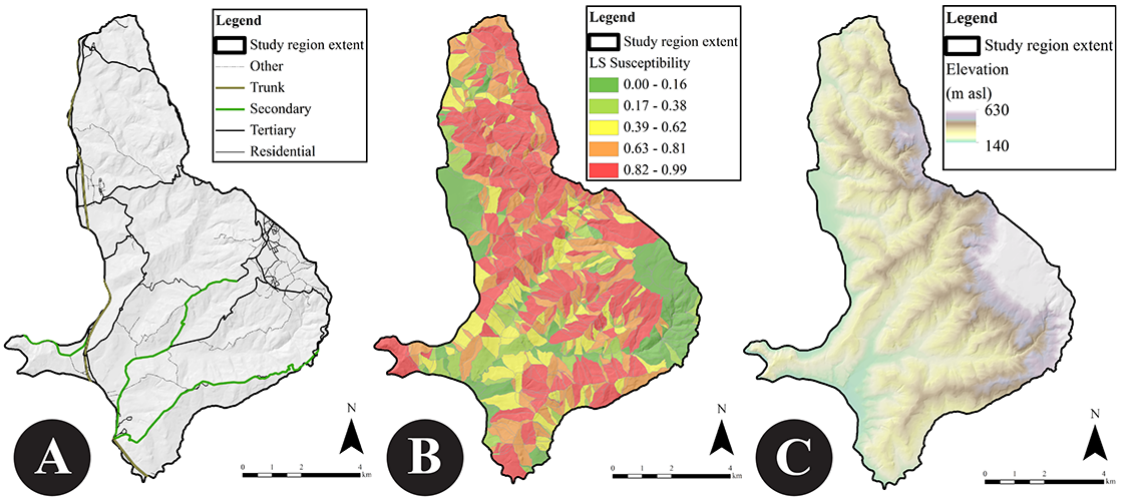


Figure 7.11 *Landslide-Road Impact Model (LRIM)* ***Inputs 1*** *to* ***4*** *data for the Collazzone region (Umbria, Central Italy).* ***(A)*** *Map of major and minor roads (OSM, 2014) (****Input 1****),* ***(B)*** *Landslide susceptibility map at slope unit scale (Rossi et al., 2010) (****Input 2****),* ***(C)*** *Digital elevation model (10 m × 10 m spatial resolution) used to calculate slope aspect (Ardizzone et al., 2007) (****Input 4****). Boundary of the maps in (A) to (C) constitutes the study area boundary (****Input 4****). The scale shown is 4 km.*

Each of the **Inputs 1** to **6** is now discussed in more detail.

7.3.2.1 Road Map (Input 1)

A vector map of the road network of the region is required to identify where the synthetic triggered landslide event inventories generated intersect with the roads. Road maps are available from a variety of sources, but can vary greatly in terms of:

- Mapping quality

- How current the information is
- Level of detail
- Homogeneity of coverage
- Who has access to the data

For the LRIM process, road network maps that include major and minor roads and that have a high level of spatial accuracy (so as to align with other data) are used. Some road maps with global coverage are freely available (e.g., Digital chart of the world, Open Street Map), thus allowing the model to be run at low cost, although they vary in terms of 'quality' (outlined in the above bullet points). **Figure 7.12** illustrates some road maps for the Collazzone region (Umbria, Central Italy) region where several different type road network maps are available that vary in the degree to which they meet the criteria listed above:

- *Digital chart of the world* (Danko, 1992) (**Figure 7.12A**) is a freely available global level map of major roads, originally created to help pilots with medium altitude navigation, last updated in 1992. Although global in coverage, the age and quality of data does vary regionally, and spatial accuracy is generally low.
- *The Italian Government Map Portal* (GN, 2014) (**Figure 7.12B**) is produced by the Ministry of the Environment at 1:10,000 scale (GN, 2014) although it is not clear whether this data is freely available to the public.
- *Open Street Map (OSM, 2014)* (**Figure 7.12C**) data is a global open source, freely available portal for maps. Maps are created from a combination of public domain maps and volunteer mapping, so quality and detail are not homogeneous at a regional level (Haklay and Weber, 2008). For Collazzone, visual comparison with other available maps suggests this source has relatively good coverage.
- *DB Prior* (**Figure 7.12D**) is a national scale pool of regional maps, corrected for network connectivity (CISIS, 2014). This map contains the major roads of the region, but excludes urban streets.
- *Detailed road map created by the CNR-IRPI from aerial photography* (Ardizzone *et al.*, 2008) (**Figure 7.12E**). This map has the benefits of roads being mapped as areas (rather than lines), having a fine scale of detail and being from the same date as the triggered landslide event which is simulated in the model. However, maps of this quality are rarely freely available due to the time involved in their creation. This map is used for data exploration and model confrontation purposes but acknowledge this is an unrealistically high quality source of data for most regions where the LRIM will be applied.

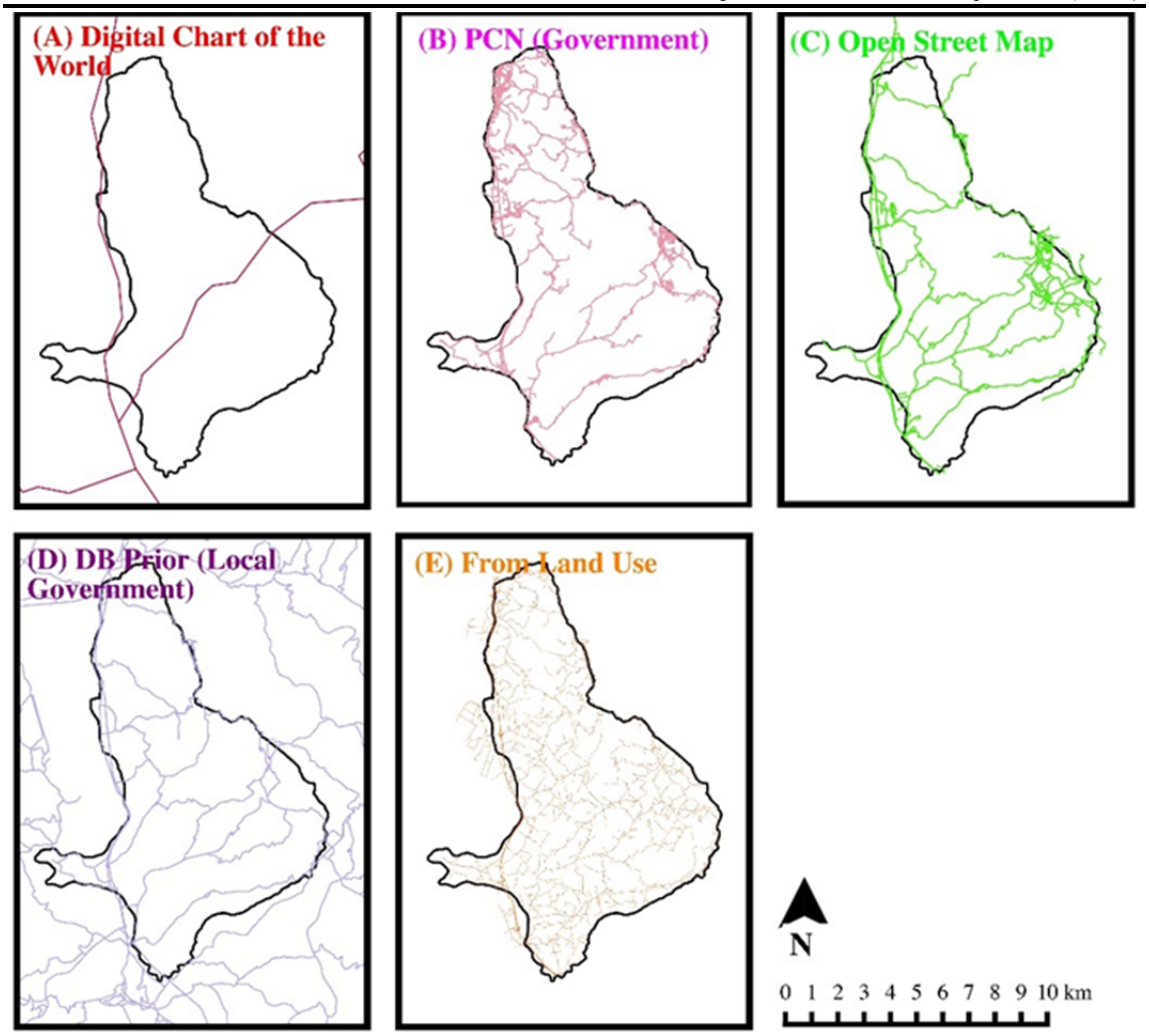


Figure 7.12 *Various road network maps available for the Collazzone study region.* ***(A)*** *Digital chart of the world (Danko, 1992).* ***(B)*** *National Government Map data (Geoportale Nazionale) produced by the Ministry of the Environment (GN, 2014).* ***(C)*** *Open Street Map (OSM, 2014).* ***(D)*** *DB Prior which is a national scale pool of regional maps (CISIS, 2014).* ***(E)*** *Detailed road map created by the CNR-IRPI from aerial photography. The scale shown is 10 km.*

From visual comparison of the road maps (**Figure 7.12**) there are distinct differences in coverage and data quality. This is likely to be the case for many other regions of the world where the LRIM can be applied. Maps **B**, **C** and **E** in **Figure 7.12** are suitable for use in the LRIM. Of maps that are freely available for all countries of the world (Digital Chart of the World, **Figure 7.12A** and Open Street Map, **Figure 7.12C**), Open Street Map (OSM) is often more suitable in terms of level of detail that roads are mapped to, and we would recommend this as the principal source of road map data in countries where 'official' maps are not available. It is a known issue that OSM coverage tends to be more detailed in urban areas, where there are more registered users (Mooney and Corcoran, 2014), which could potentially present problems when applying the LRIM to remote mountainous areas. Nonetheless, others have noted that OSM coverage is often better than that of National Mapping Agencies in many developing countries (De Leeuw *et al.*, 2011), and several large

commercial companies now use OSM data for products and services (Mooney and Corcoran, 2014), adding confidence to the use of OSM data in the LRIM model. Moreover, in recent years, there have been 'drives' to complete maps in the run up to hazard events, such as the Humanitarian Open Street Map Team (HOT), who mobilised to supplement existing maps of the Philippines in the days before Typhoon Haiyan made landfall in 2013 (Butler, 2013). These networks of 'digital humanitarians' are becoming increasingly formalised and coordinated, and could potentially be a future source of not only high quality road network maps, but also landslide inventories and impact data in the coming years.

7.3.2.2 Landslide Susceptibility Map (Input 2)

Input 2 is a landslide susceptibility map (**Figure 7.11B**) which partitions a region into zones of differing probability of landslide spatial occurrence under a given set of geo-environmental conditions (Guzzetti *et al.*, 2006b). The landslide susceptibility map is used to condition where more or less landslides are dropped in the LRIM (**Method 6**). For two out of three of our study regions (Collazzone region, Umbria, Central Italy and Oat Mountain Quadrangle, Northridge, California, USA, discussed later in **Section 7.4**), pre-existing landslide susceptibility maps were available. If a susceptibility map is not available, it may be possible for the user to create an 'approximation' (potentially un-validated) of susceptibility using a Digital Elevation Model (DEM), a small number of soil parameters and freely available tools such as SHALSTAB (Montgomery and Dietrich, 1994). Coe *et al.*, 2004; Guinau *et al.*, 2007) suggested that landslide density from an inventory (Guzzetti, personal communication, May 2014; Fell *et al.*, 2008) may be reasonable proxies to determine where more or less landslides are distributed in the LRIM model. Certainly, these heuristic techniques may be subject to more error than a typical landslide susceptibility map, but allow the application of LRIM in data-poor scenarios. Indeed, it is possible to run LRIM without a landslide susceptibility map, as the locations of landslides are randomly generated in the LRIM model.

7.3.2.3 Study Area Boundary (Input 3)

Input 3 is a vector map of the study area boundary, which is visible in **Figure 7.11A,B,C**. Due to the way (*N, E*) locations are generated for landslide points (**Method 3**), if the study area is not rectangular, some (*N, E*) locations may be generated outside the study area. The study area boundary is used to reject any points falling outside the region of interest.

7.3.2.4 Digital Elevation Model (Input 4)

Input 4 is a Digital Elevation Model (**Figure 7.11C**) which is used to calculate a slope aspect map (**Method 2**), which is then used in **Method 7** to align the long axis of landslides in the downslope direction to make the synthetic landslide inventory map more realistic.

7.3.2.5 Landslide Density (L_D) (Input 5)

Input 5 is the landslide density (L_D) where the user defines the average number of landslides per square kilometre (L_D) to be dropped over the study region (although the actual landslide density will vary over the study region due to landslide susceptibility). The area of the study area boundary (**Input 3**) and landslide density (**Input 5**) are used to calculate the number of landslide point locations to generate (**Method 3**).

By varying L_D, the user can simulate different triggered landslide event scenarios. For instance, following a triggered landslide event, if the number of landslides can be rapidly approximated using techniques such as high temporal (low spatial) resolution remote sensing (e.g., Mota *et al.*, 2014), LRIM can be used to hindcast the magnitude of the event (number of landslides) and forecast the impact of that event in near-real time.

7.3.2.6 Broad Characteristics of Road Landslide Protection in Region (Input 6)

From initial exploration of LRIM output (**Section 7.5**), we found that there may be interactions between different types of roads and landslide susceptibility that were not captured at the scale of the landslide susceptibility maps used. Indeed, **Section 7.2.2** (*Roads and Landslide Susceptibility*) described a range of studies showing that landslide incidence increased with increasing proximity to roads (e.g., Brenning *et al.*, 2015).

Input 6 is the broad characteristics of road landslide susceptibility due to protection measures (or lack of) in a given region, and includes also the increased vulnerability to landslides that might occur due to poor placement or installation of a road. For instance, the corridor along a road may be *less susceptible* to landslides than the surrounding area if road protection measures have been integrated into the construction (Fookes *et al.*, 1985). Some key types of road protection measure are listed in **Table 7.3**. Maps of road landslide protection measures are difficult to identify as various organisations may be involved in their installation and maintenance over long periods of time (e.g., highways agencies, local government, private landowners) (Dashwood, personal communication, 2014; Klose, personal communication, 2015).

Table 7.3 *Commonly used protection measures on landslide-prone roads.*

Protection Measure	Description/ Purpose	Reference	Example Image
Barriers	Designed to trap material and absorb energy to either stop or slow landslide movement.	Geobrugg (2012)	
Bridges	Reduce risk of flow-type landslides intersecting the road by increasing road height.	Petley (2011)	
Drainage	Improved drainage can increase friction and cohesion and reduce self-weight of the slope. Can also reduce erosion.	Gerscovitch and Costa (2004) Image: ICEA (2012)	
Monitoring Equipment and Alert Signs	If landslide precursors (e.g. a rainfall threshold or movement) are detected by the remote monitoring equipment, road side warning systems are activated.	Transport Scotland (2012)	

Protection Measure	Description/ Purpose	Reference	Example Image
Slope Reinforce-ment	E.g. Applying concrete (shotcrete), inserting reinforced bars into the slope to increase stability.	Tan and Chow (2004) Image: IECA (2012)	
Slope Stabilisation Nets/ Meshes/ Matting	May trap/reduce velocity of small movements, promote vegetation growth and reduce erosion.	Ortigao (2004). Image: Roctek International (n/d)	
Tunnels	Landslide moves over tunnel roof rather than over the road. Can reduce or remove the risk entirely.	WSL (2011)	
Vegetation	Controls erosion, can strengthen soil and catch or reduce velocity of small landslides.	Gerscovitch and Costa (2004). Image: Geobrugg (2012)	

A road may alternatively *increase* the likelihood of a landslide occurring by changing the structure and drainage of a slope (Larsen and Parks, 1997). The spatial resolution of data used to create landslide susceptibility maps may not capture these landslide-road interactions. To attempt to simulate these landslide-road interactions, three LRIM modes of operation have been developed for the susceptibility in the corridors along roads (relative to the susceptibility for the local area in which the road occurs):

- **Mode A** (Standard Susceptibility): landslide susceptibility is neither increased nor decreased in the corridor along roads (i.e., it is left the same as originally given in the susceptibility map).
- **Mode B** (Increased Susceptibility): landslide susceptibility is increased in the corridor along roads (relative to the susceptibility for the local area in which the road occurs).
- **Mode C** (Decreased Susceptibility): landslide susceptibility is decreased in the corridor along roads (relative to the susceptibility for the local area in which the road occurs).

For **Mode B** (increased road landslide susceptibility) and **Mode C** (decreased road landslide susceptibility) the landslide susceptibility map (**Input 2**) is adjusted along road corridors by randomly selecting portions of the road network, creating a buffer of a predetermined width and either reducing or increasing landslide susceptibility by a predetermined amount. Some potential values for the width of these buffers were derived by examining the linear distance between landslides and the nearest road (by road type) for three observed triggered landslide event inventories and road network maps we have available, which will later be described in detail in **Section 7.4**. These inventories are:

- *Collazzone*. 413 landslides triggered by snowmelt in 1997 in Collazzone, Central Italy (Cardinali *et al.*, 2000).
- *Northridge*. 11,111 landslides triggered by the 1994 Northridge Earthquake in California, USA (Harp and Jibson, 1995).
- *Su-Hua*. 149 landslides triggered by the 2010 Typhoon Megi along the Su-Hua road corridor, Northeast Taiwan (Chen, personal communication).

For each inventory, the mean centroid of each landslide polygon and the distance to the nearest road (*D*) was calculated in a GIS. An illustration of this process is shown in **Figure 7.13**. *D* was then separated by the type of road the landslide was nearest to (primary, secondary, tertiary or minor). **Figure 7.**14 shows violin plots of *D* values by road type for each of the three inventories above. Violin plots show a boxplot of D on the *y* axis, and

also kernel density (i.e., where more or less of the data points lie) on the *x* axis (Hintze and Nelson, 1998).

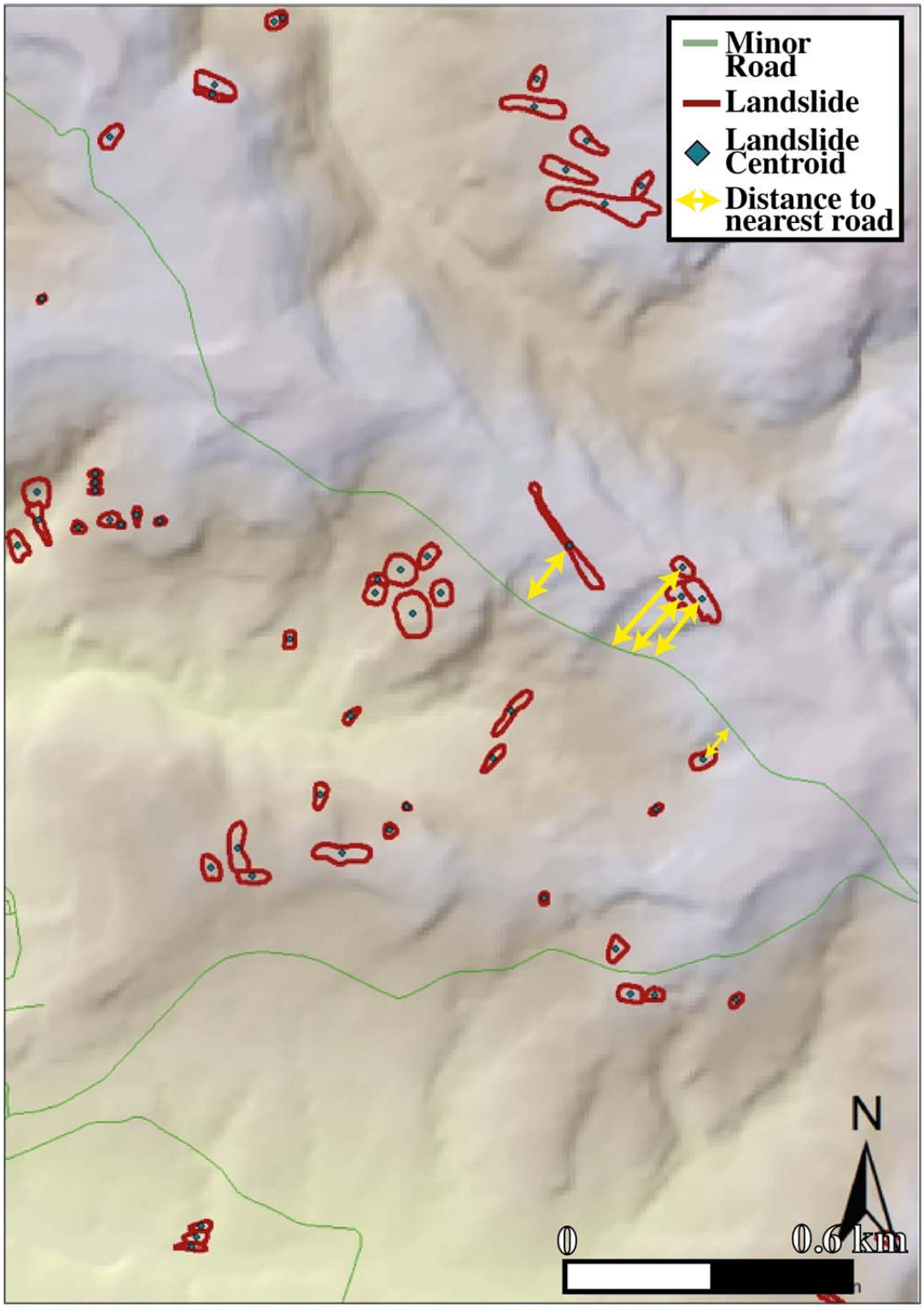


Figure 7.13 *Calculation of the minimum linear distance between observed landslide centroids and the nearest road (D) for subset of landslides triggered by 1997 snowmelt in Collazzone, Central Italy. The yellow arrows illustrate some linear distances between landslides and the minor road (green line) running through the area. Distances (D) are calculated for all landslide centroids to the nearest road, and the type of that road recorded.*

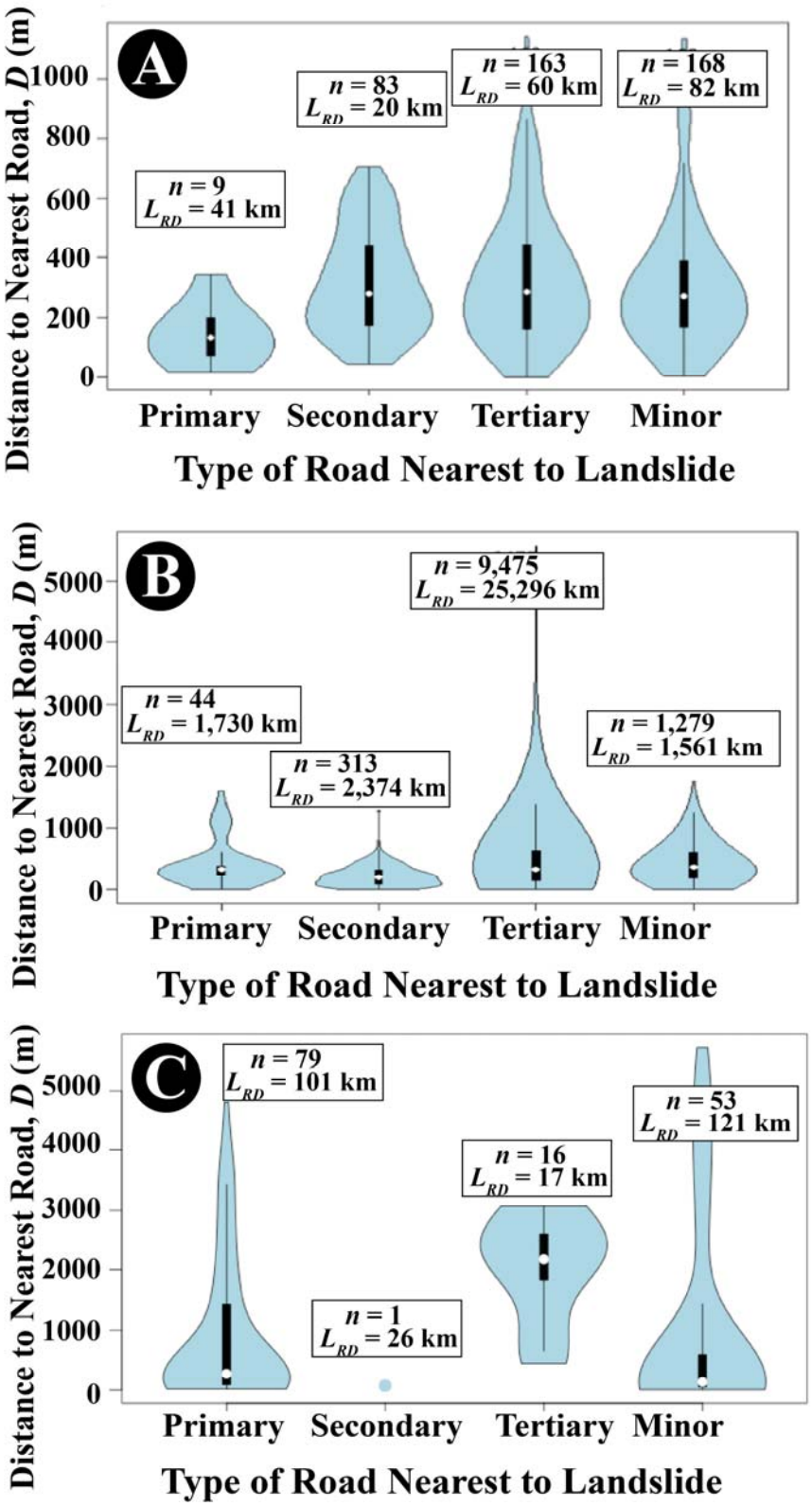


Figure 7.14 *Violin plots of distances between landslide centroids and the nearest road, split by road type for* ***(A)*** *413 landslides triggered by 1997 snowmelt in Collazzone, Central Italy* ***(B)*** *11,111 landslides triggered by the 1994 Northridge Earthquake, California, USA* ***(C)*** *149 landslides triggered by 2010 Typhoon Megi in the Su-Hua area, Northeast Taiwan. Note: only one distance is calculated for each landslide centroid, to the nearest road, and the type of that road recorded. Each inventory, including sources of data is described further in* ***Section 7.4****.* L_{RD} *corresponds to the total length of that road type across the region, n corresponds to the number of landslides closest to that type of road.*

Figure 7.14 illustrates some similar behaviour in landslide-road interactions between Collazzone, Italy (**Figure 7.14A**) and Northridge, USA (**Figure 7.14B**):

- *The nearest type of road* to most landslides in each inventory is tertiary or minor.
- *The probability density of observing small values of the distance between landslides and nearest roads (D)* is relatively low (i.e., there are less landslides very close to roads, and more at middle distances to roads, illustrated by the tapering of the violin plots in the *x*-axis for small values of *D*).
- *The variability in distance to nearest road (D) for tertiary and minor roads is high*, but we attribute this to more observations in these categories, and a greater length of minor roads than major roads in these regions.

The landslide-road interactions for Su-Hua (**Figure 7.14C**) are less clear, but suggest a greater probability of observing landslides at small distances from the road (illustrated by the 'bulbous' base of the violin plots), and that landslides are often closest to primary roads. These results for Su-Hua, Taiwan, possibly confirm the increase in landslide occurrence with increasing proximity to roads as found by others for unimproved roads and lower-income countries (see **Section 7.2.2** *Roads and Landslide Susceptibility*). However, both the landslide inventory and road network map used for this location are lower confidence in terms of completeness, and potential bias towards landslides that have occurred close to the road.

In LRIM **Mode C** (decreased landslide susceptibility in proximity to roads), the second percentile of *D* values (distance from landslides to nearest roads) for each road type is used as the width for the road buffer. Within this road buffer zone, landslide susceptibility (S_{LS}) is reduced to $S_{LS} = 0$. The proportion of the road network by road type selected to randomly buffer is based on the percentage of landslides within the inventory that are closest to that type of road. For example, in Collazzone, 2 % of the observed inventoried landslides were closest to major roads, so 98 % of the road network is buffered. The specific values used for each study site are presented later in **Section 7.4**.

LRIM road buffering is demonstrated in **Figure 7.15** where two short segments of road have been buffered and landslide susceptibility reduced from ($0.80 < S_{LS} \leq 0.99$) to ($0.00 < S_{LS} \leq 0.20$). In **Mode B** (increased road landslide susceptibility), a greater proportion of minor road length will be buffered and susceptibility increased. This part of

the model could be developed further to derive more robust buffer widths and susceptibility changes.

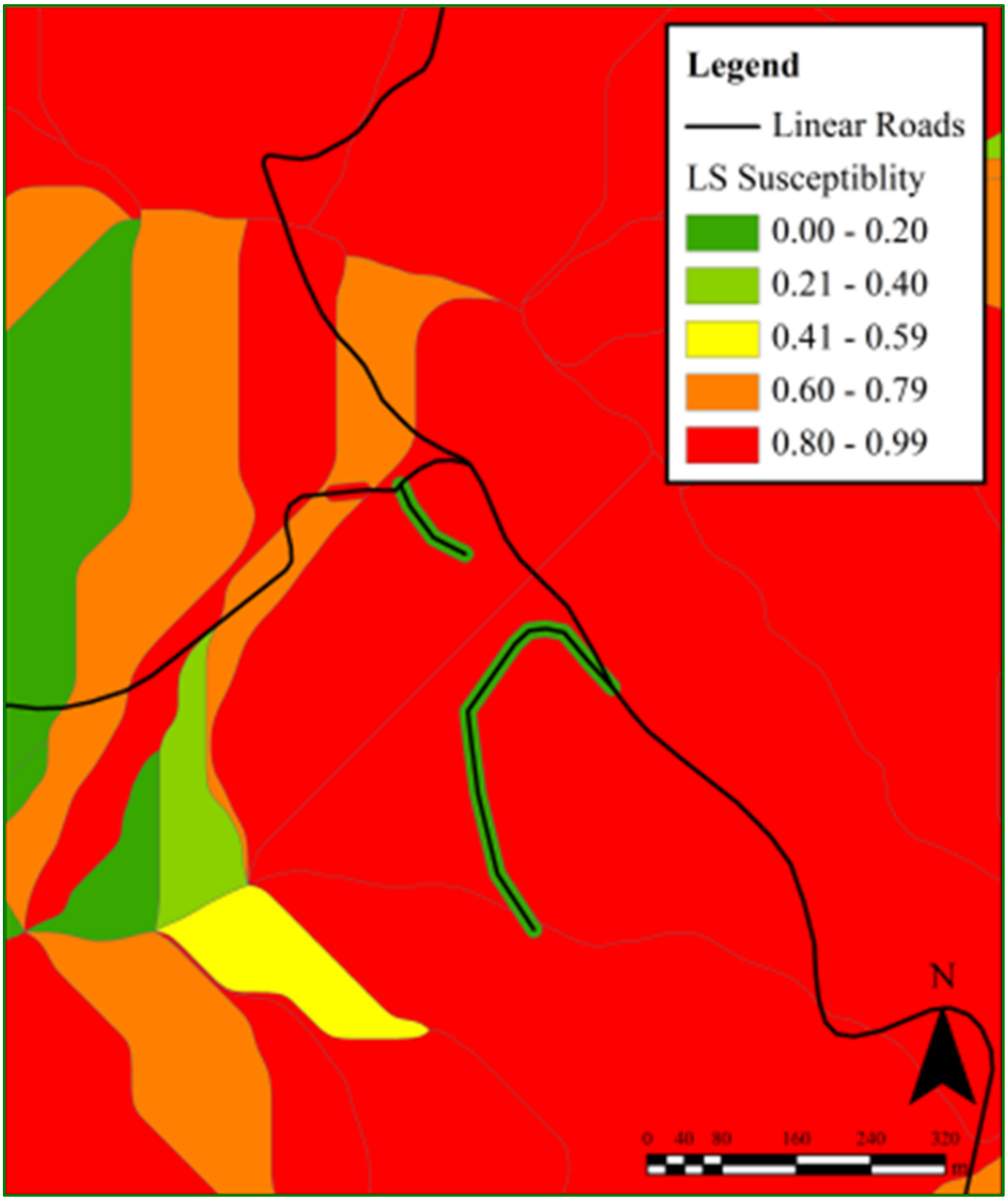


Figure 7.15 *Road Landslide Susceptibility Mode B (reduction of susceptibility). An example of random selection of road portions that are buffered and landslide susceptibility (S_{LS}) adjusted within this buffer zone. In this example, two short lengths of road have been randomly selected and buffered by 20 m either side of the road line, for a total buffer zone of 40 m. Within this buffer zone, landslide susceptibility has been reduced from S_{LS} = 0.9 to S_{LS} = 0.0. This means that no landslide centroid will occur within these buffer zones. The scale shown is 320 m.*

7.3.3 LRIM Model Pre-loop (Methods 1 to 2)

After the user has input data and set the landslide density and LRIM mode of operation (**Section 7.3.2**), some initial processing is performed before the Monte-Carlo loop (**Methods 3** to **9**) is commenced:

7.3.3.1 Calculation of Number of Landslides to Drop (N_{LS}) (Method 1)

The areal extent of the study area boundary (**Input 3**) and landslide density (L_D) (**Input 5**) are used to calculate the number of landslide point locations (N_{LS}) to drop. Because many of these point locations are rejected later in **Method 6**, $n = 10 \times N_{LS}$ are generated, where n represents the number of points generated internally by LRIM, and N_{LS} are the total number of landslides in the final triggered landslide event synthetic map.

7.3.3.2 Create Aspect Map (Method 2)

Two different slope aspects maps are created. The first slope map is used to align the landslide ellipse long axis with the slope direction (**Method 6**). This aspect map is calculated from a DEM (**Input 4**) using the standard GRASS GIS tool (r.slope.aspect), which includes a 3 pixel × 3 pixel window to calculate the slope direction of each pixel (Shapiro and Waupotitsch, 2011) (**Figure 7.16A**). Due to small scale variations in slope aspect, a coarse scale measurement of aspect is required when aligning the travel direction of the landslide with the general aspect of the slope. For LRIM, a DEM of 100 m resolution is used.

The second slope aspect map is at the slope unit scale and will be used later in **Method 8** to prevent landslides from crossing ridges or drainage divides (and thus unrealistically travelling downhill and then back uphill). This slope unit scale aspect map has one value of average azimuth for each slope unit (**Figure 7.16B**). Many landslide susceptibility maps are produced at the slope unit scale, so these divisions can be used in the model. If a map of slope unit divisions is not available (e.g., if the landslide susceptibility map is in a raster format), the user can create one using an already existing web processing tool, a DEM, and an optional map of alluvial zones as described by Alvioli *et al.* (2014), shown in **Figure 7.31C**. The average value of azimuth of all raster cells within each slope unit polygon can then be calculated using the GRASS GIS v.rast.stats tool.

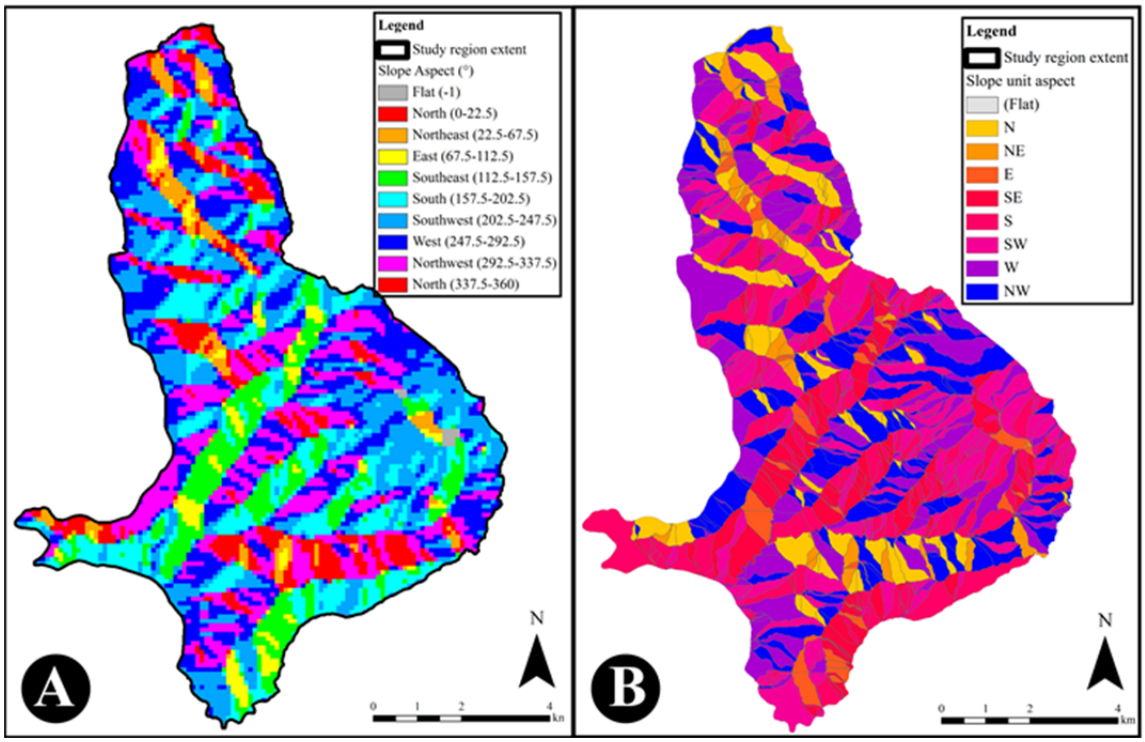


Figure 7.16 *Maps of slope aspect for the Collazzone region (Umbria, Central Italy).* ***(A)*** *Slope aspect calculated from a coarse spatial resolution DEM.* ***(B)*** *Slope aspect aggregated to slope unit level. The scale shown is 4 km.*

7.3.4 LRIM Model Synthetic Landslide Attributes (Methods 3 to 5)

7.3.4.1 Random Generation of (*N, E*) Point Locations (Method 3)

A total of $n = 10 \times N_{LS}$ random (N, E) points are generated by selecting values from uniform distributions between the minimum and maximum Northings and Eastings of the study area boundary. Because the study area boundary may not be rectangular, some of these points may lie outside the area of interest and are rejected later in **Method 6**. For each (N, E) point generated in **Method 3**, a random uniform value $0 \leq x \leq 1$ is also generated. This will also be used in **Method 6** in conjunction with the landslide susceptibility map (**Input 2**) to ensure more landslides occur in high susceptibility zones compared to low susceptibility zones.

7.3.4.2 Generation of Landslide Areas (Method 4)

In this step landslide areas are randomly sampled from known frequency-area distributions (in other words, the relative proportion of large, medium, small, very small landslides we would like in our final synthetic landslide inventory maps). The behaviour of triggered landslide events in terms of frequency-area statistics appears to be similar for low mobility landslides in medium-high topographic zones, regardless of location, magnitude of the event and triggering mechanism (Stark and Hovius, 2001; Malamud *et al.*, 2004a) (**Chapter**

4). Although different authors use different probability density functions to describe this behaviour (Van Den Eeckhaut *et al.*, 2007), the overall frequency-size distribution of low-mobility landslide areas is similar: the probability density is relatively low for very small landslides (e.g., $A_L \approx 10$ m^2) and increases with increasing landslide area (A_L) until the distribution peaks at landslide areas of $A_L \approx 400$ m^2. After this maximum probability (where there are the highest number of landslides) the frequency (probability) of landslides decreases as landslide areas increase, as an inverse power law function. One example of a probability density function fit (discussed previously in **Chapter 4**, but re-introduced here) to three substantially complete different large triggered landslide event inventories is the three parameter inverse-gamma distribution (Malamud *et al.*, 2004a):

$$p(A_L \mid \rho, a, s) = \frac{1}{a\Gamma(\rho)}\left[\frac{a}{A_L - s}\right]^{\rho+1} \exp\left[-\frac{a}{A_L - s}\right] \qquad \textbf{(7.1)}$$

with $\Gamma(\rho)$ the gamma function of ρ, parameter a controlling the exponent of the inverse power law, i.e. right tail, s controlling the amount the left tail bends as part of the exponential, and a primarily controlling the position of the rollover. The inverse-gamma distribution is shown in **Figure 7.17** and described in more detail in **Chapter 4**. For three large substantially complete triggered landslide event inventories shown in **Figure 7.17**, Malamud *et al.* (2004a) found a general inverse-gamma distribution that applied to the landslide areas in all three events, with parameters (**Eq. 7.1**), $\rho = 1.40$, $a = 1.28 \times 10^{-3}$ km^2, $s = -1.32 \times 10^{-4}$ km^2.

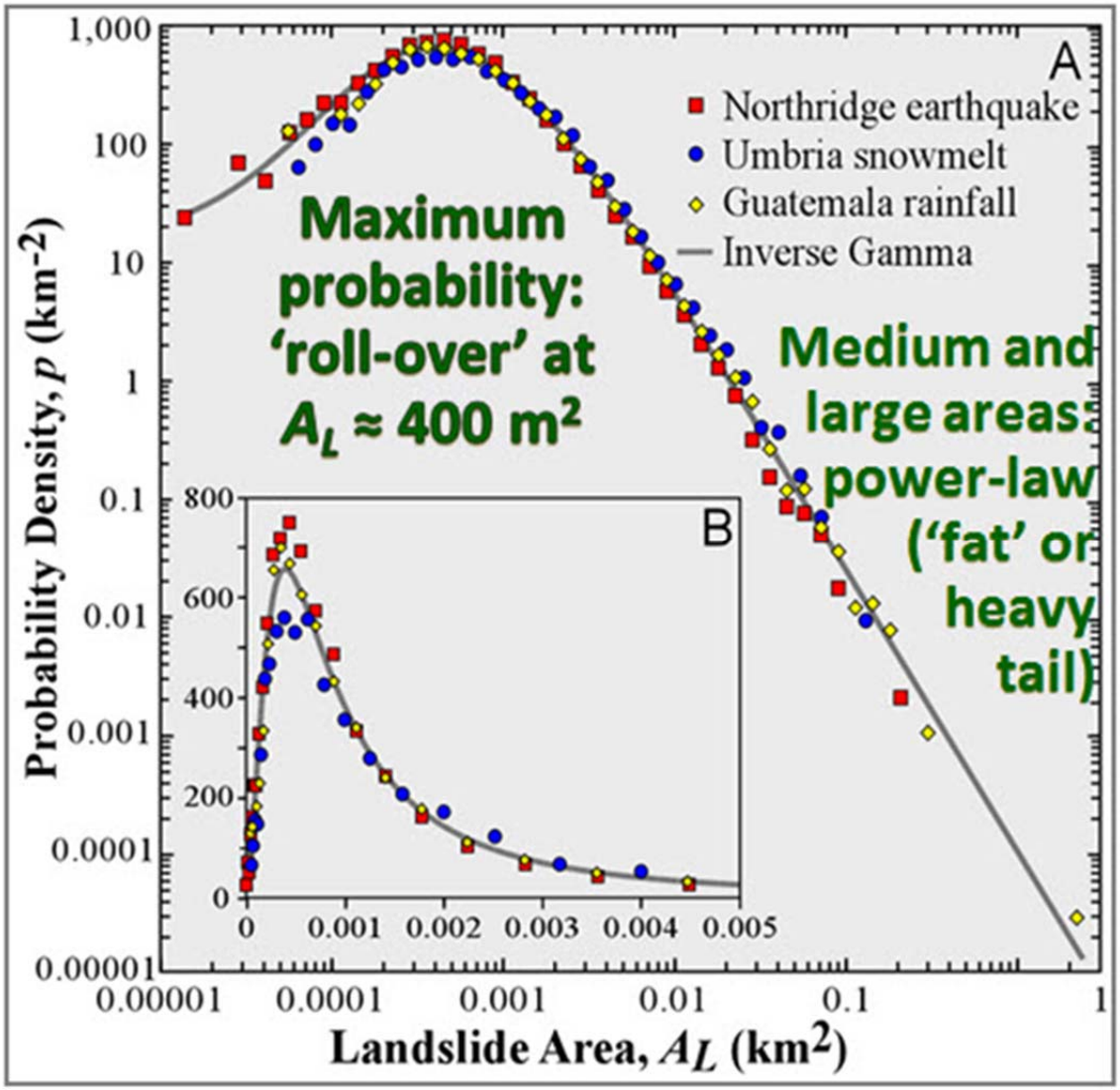


Figure 7.17 *Inverse-gamma probability density function fit to frequency densities from three substantially complete triggered event inventories (Malamud et al., 2004a).*

A probability distribution function (pdf) can be used to create a synthetic inventory of individual landslide areas by randomly selecting values from the pdf. In the LRIM, landslide areas are randomly selected from the inverse-gamma pdf (**Eq. 7.1**) using the inverse transform sampling method (Press *et al.*, 2002), executed in the Python Scipy code (Jones *et al.*, 2014). The following steps are applied (this is also illustrated in **Figure 7.18**):

For each value of A_L to be randomly selected from the pdf:

- Generate a random uniform value, $0 \leq x \leq 1$.
- Take the inverse value of the cumulative distribution (F) where $F(A_L) = x$.
- This inverse value is a randomly generated value of A_L which follows the inverse-gamma probability density function.

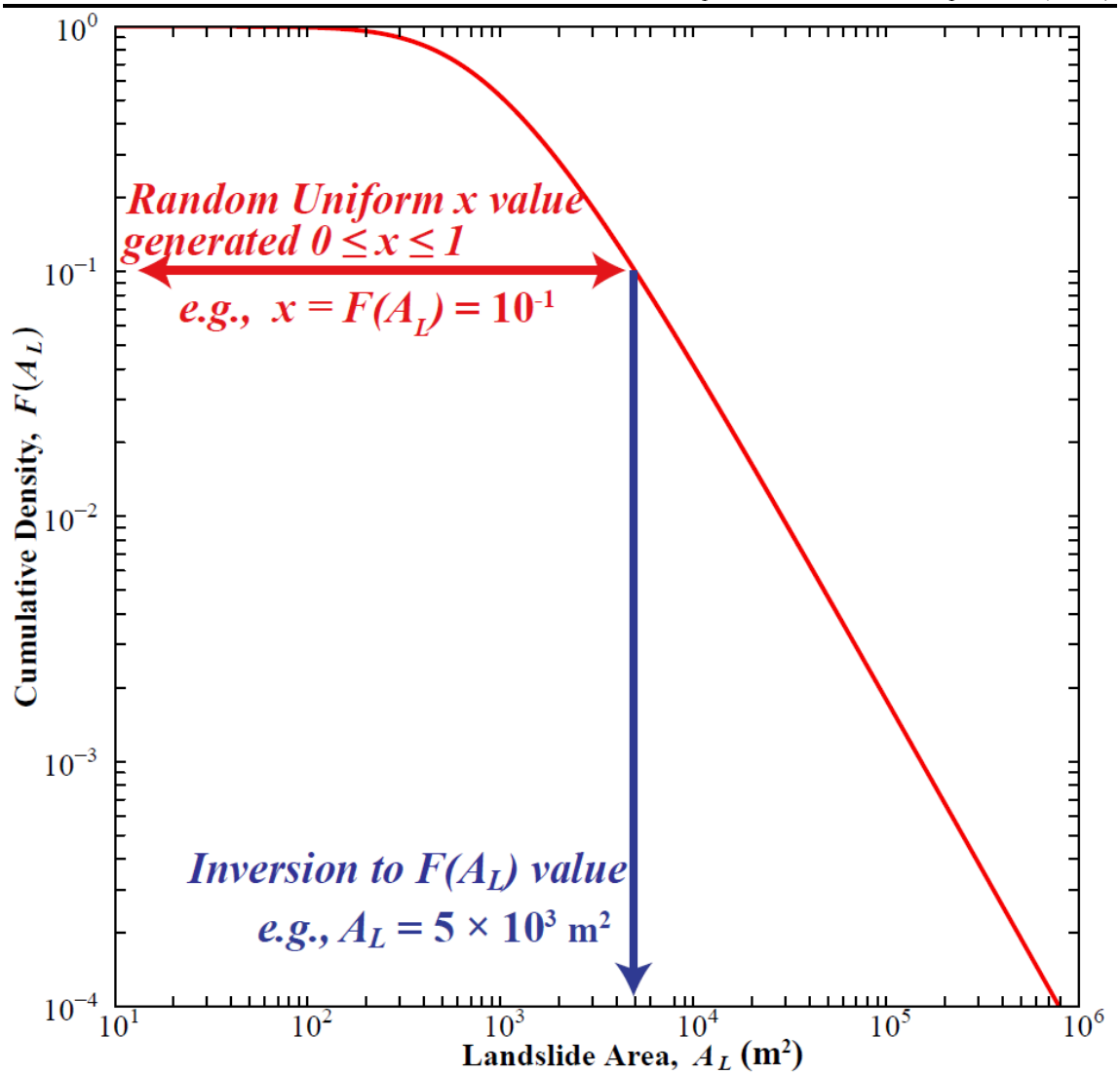


Figure 7.18 *Demonstration of the inversion method used to randomly select landslide areas (A_L) from the three parameter inverse-gamma pdf (i.e., the population), whilst preserving the shape of the statistical distribution within the sample. A random uniform value $0 \leq x \leq 1$ is generated, where $x = F(A_L)$ (cumulative probability), and from this the corresponding random landslide area (A_L) value chosen. This process is repeated such that the sample has the number of desired landslide areas (A_L), N_{LS}, which will follow the inverse-gamma pdf.*

By repeatedly sampling from the same statistical distribution, a synthetic inventory of N_{LS} landslide areas is generated. Although the absolute values randomly selected from distribution will vary, the overall frequency-size distribution should be roughly the same as the original inverse gamma probability density function the values are selected from. This is illustrated in **Figure 7.19**, showing three different sets of samples from the inverse gamma pdf with parameter values from Malamud *et al.* (2004a). Each set of samples differs in the number of random values generated; as more values are selected, the more the distribution of probability densities approaches the original distribution the values were generated from.

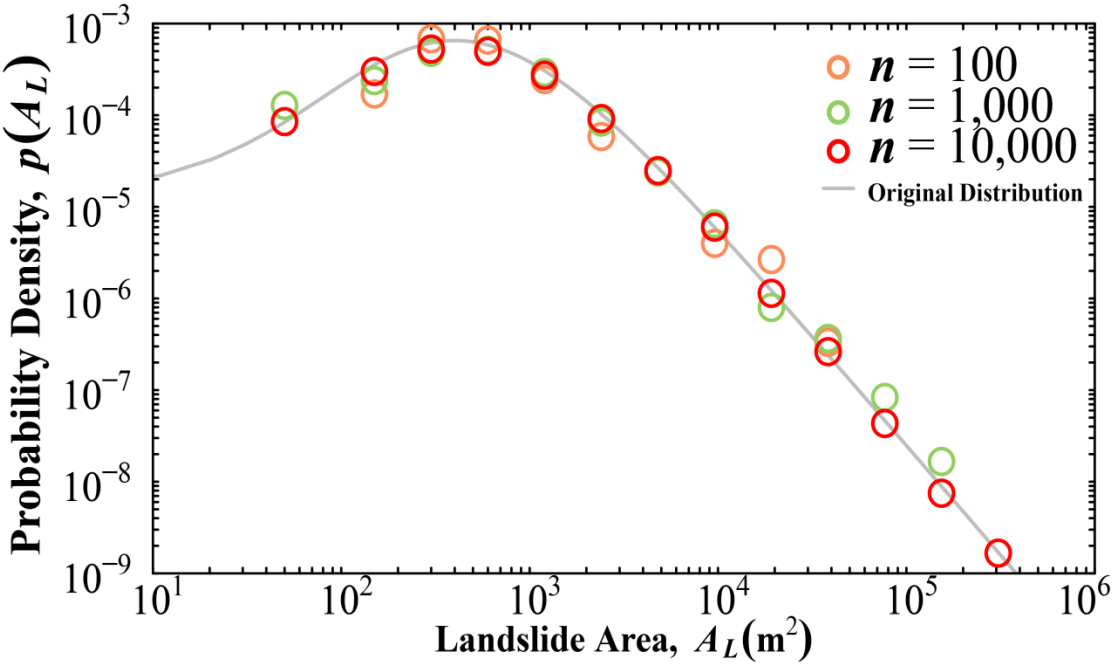


Figure 7.19 *Examples of probability densities of landslide area for three different 'sets' of landslide areas randomly selected from the inverse gamma probability density function (**Eq. 7.1**) with parameter values from Malamud et al. (2004). The original inverse gamma pdf the samples are drawn from is shown as a grey line. Probability densities were calculated using approximately logarithmically increasing bin sizes. As the size of the 'set' increases (i.e., we randomly select more values of landslide area from the pdf), the set more closely resembles the original distribution.*

In the LRIM, it is possible to use other statistical distributions that have been used to describe landslide area probability density functions, such as the following:

- *Double Pareto pdf* (Stark and Hovius, 2001).
- *Power-law distributions to model triggered rockfall events* (e.g., Hungr *et al.*, 1999; Guzzetti *et al.*, 2004b; Malamud *et al.*, 2004a; Barlow *et al.*, 2012).
- *Inverse-gamma pdfs with other parameter values* established by using a freely available landslide statistics software LStats (Rossi *et al.*, 2012; Rossi *et al.*, 2014) after that software tool has been used to find the best fit statistical distributions to other inventories.

However, in the LRIM, the parameters are set to those as used by Malamud *et al.* (2004a) as a 'general' landslide distribution, for which to choose our landslide frequency-area statistical distribution.

7.3.4.3 Generation of Landslide Elliptical Shapes (Method 5)

By using **Methods 3** and **4** LRIM has created a table of n random (N, E) locations which each have an associated landslide area (A_L) where the overall distribution of landslide areas follows an inverse-gamma probability density function with the parameters as given by

Malamud *et al.* (2004a). In order to map these landslide areas, each must be given a shape (and later, discussed in the next section, an azimuth).

Primary and secondary research (outlined in more detail in **Chapter 5**) was undertaken to investigate (i) what shape landslides are typically and (ii) does this shape vary with size. Several references were found to support the claim that an ellipse is a reasonable approximation of most landslide shapes, although the dimensions of these ellipses can vary (e.g., Goltz, 1996; Marchesini *et al.*, 2009). In order to investigate how much the dimensions of landslide ellipses vary, and how close actual landslides are to an ellipse, we initially used two substantially complete large triggered landslide event inventories to create best fit elliptical approximations of each landslide polygon, calculate the goodness of fit of the ellipse to these landslides and then calculate the length to width ratio (L/W) of each ellipse. The inventories are (i) 11,111 landslides triggered by the 1994 Northridge Earthquake in California, USA (Harp and Jibson, 1995) and (ii) 9,594 landslides triggered by Hurricane Mitch in Guatemala (Bucknam *et al.,* 2001). We found that an ellipse is a reasonable approximation of 82–85 % of landslides within each inventory.

For each ellipse accepted as a reasonable fit, the ratio of the long axis to short axis was calculated. This is the length (L) to width (W) ratio of the landslide, L/W. For each inventory, probability densities of L/W were calculated and separated into categories based on landslide area (A_l), where category size increased approximately logarithmically. We experimented with fitting various probability density functions to this data using Maximum Likelihood Estimation (MLE) (Myung, 2003). An inverse-gamma probability density function was found to be a statistically significant fit to the probability density distribution of L/W in each landslide area category within each inventory (tested using a Monte Carlo Kolmogorov-Smirnov goodness-of-fit test, described by Clauset *et al.*, 2009), shown in **Figure 7.20**.

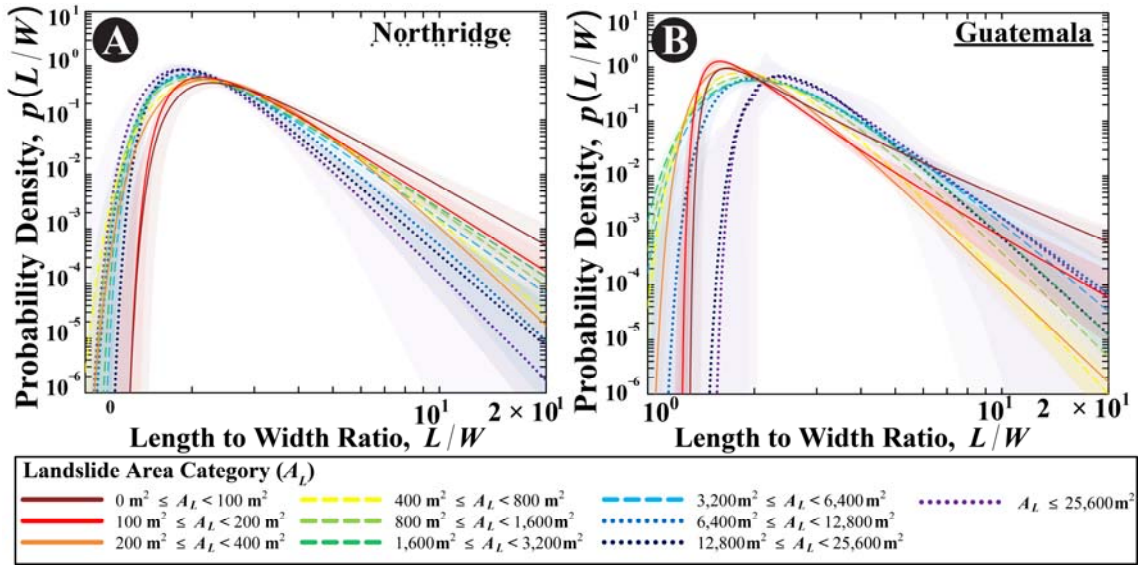


Figure 7.20 *Inverse Gamma Probability Density Functions (pdfs) fit to landslide ellipse length to width ratio (L/W) values in each landslide area category.* ***(A)*** *9,441 landslides in the Northridge inventory and* ***(B)*** *8,031 landslides in the Guatemala inventory. The shaded area either side of the pdf line represents a form of 5th/95th percentile confidence intervals around the pdf. These are calculated using a bootstrapping technique (**Section 5.6.2.2**) where observed data is repeatedly sampled with replacement and an inverse gamma pdf fit to each sample.*

In the Northridge inventory (**Figure 7.20A)**, we observe that as landslide area category (A_L) increases, the distribution shifts towards smaller values of L/W and the gradient of both tails broadly increases. Broadly speaking, this implies that in Northridge, larger landslide areas tend towards more compact shapes with smaller values of L/W. The opposite is observed in the Guatemala inventory (**Figure 7.20B**); as A_L increases, the inverse gamma pdf shifts to larger values of L/W.

Chapter 5 showed the application of this methodology to five additional smaller, lower confidence inventories. For each landslide area category within each inventory, an inverse gamma pdf was a statistically significant model of L/W probability density distribution, although there was considerable variety in the parameter values describing each pdf. Generally, landslide inventories triggered by moisture (rainfall or snowmelt) had similar overall behaviour to the Guatemala inventory (**Figure 7.20B**), and it is possible that deviation from this behaviour was more a result of small sample sizes and less complete inventories used in analysis. For the three study regions we apply the LRIM model to in this thesis, we have access to triggered landslide event inventories, and can thus use the best fit pdfs of L/W for each dataset, and thus landslide ellipse L/W values can be randomly selected from the inverse-gamma probability density function appropriate to that inventory. This is done using the inverse transform sampling method, described in **Method 4**.

If the LRIM is to be applied to other study regions where inventories are not available, users could select parameters based on the inventory that best reflects the environment they are modelling (e.g., in a region susceptible to earthquake-triggered landslides, the Northridge earthquake inventory L/W pdfs may be most suitable). As discussed in **Chapter 5**, analysis of more large, substantially complete triggered landslide event inventories is required to either (i) decide on a single set of parameter values for each landslide area category that model L/W for all locations or (ii) allow the user to select from a set of descriptive categories a set of parameter values that reflect L/W distribution for the inventory they wish to model. This is ongoing research.

At this stage, using **Methods 1** to **5** as illustrated in **Figure 7.9** (**Section 7.3.1**), the LRIM has generated a table of $n = 10 \times N_{LS}$ random (N, E) locations, each location with an associated landslide area (A_L) and landslide length to width ratios (L/W). An example of this LRIM output for 10 synthetic landslides is shown in **Table 7.4**.

Table 7.4 *An example of 10 LRIM randomly generated landslide (N, E) locations, areas and length to width ratios.*

ID	Northing *(N)*	Easting *(E)*	Area *(A_L)*	Length to Width Ratio *(L/W)*	Length *(L)*	Width *(W)*	Random Uniform *(x)*
#	*(m)*	*(m)*	*(m^2)*		*(m)*	*(m)*	
1	4748216	291755	624.98	2.22	21.04	9.46	0.15
2	4746156	292505	1637.24	9.59	70.71	7.37	0.84
3	4748422	290657	2254.79	2.94	45.90	15.64	0.05
4	4746624	292308	1102.80	2.26	28.14	12.48	0.36
5	4743185	290495	1692.41	2.24	34.74	15.51	0.30
6	4752568	290016	1919.79	2.21	36.77	16.62	0.04
7	4747574	291005	861.24	9.14	50.07	5.48	0.66
8	4746556	292196	7714.19	4.89	109.57	22.41	0.61
9	4746078	291554	1651.30	3.48	42.77	12.29	0.08
10	4754223	288719	1174.95	3.11	34.13	10.96	0.72

7.3.5 LRIM Model Synthetic Landslide Placements (Method 6)

At this stage of the LRIM, there are generated n = ten times the desired number of landslides (N_{LS}). As described now, in **Method 6** some of these landslides are rejected in order to create a final synthetic landslide inventory. This is a five stage process (see also **Figure 7.9**, **Section 7.3.1** for an overview) of the following steps (and illustrated in **Figure 7.21)**:

Method 6A: All n (N, E) points are mapped (**Figure 7.21A**)

Method 6B: Any points lying outside the study area boundary are rejected (**Figure 7.21B**)

[OPTIONAL]: At this stage, the landslide susceptibility map can be adjusted to take into account landslide road protection measures as described in **Section 7.3.2.6 Input 6**. If this optional step is included, the adjusted susceptibility map is used in **Method 6D**. If this optional step is not included, the landslide susceptibility map **Input 2** is used in **Method 6D**.

Method 6C: For each of the n (N, E) points, the value of landslide susceptibility S_{LS} (which ranges $0 \leq S_{LS} \leq 1$) at this location is queried. If the random uniform value $0 \leq x \leq 1$ generated in **Method 3** is such that $x > S_{LS}$ at this location, the point is rejected and the (N, E) location is deleted from the map (**Figure 7.21C**) (Zaitchik *et al.*, 2003). Otherwise the point is preserved.

Method 6D: At this stage in the LRIM, if the number of points remaining after **Method 6C** is such that $n > N_{LS}$, a randomly selected set of those points where $n = N_{LS}$ is taken using the GRASS GIS v.extract tool, with option of random selection (**Figure 7.21D**).

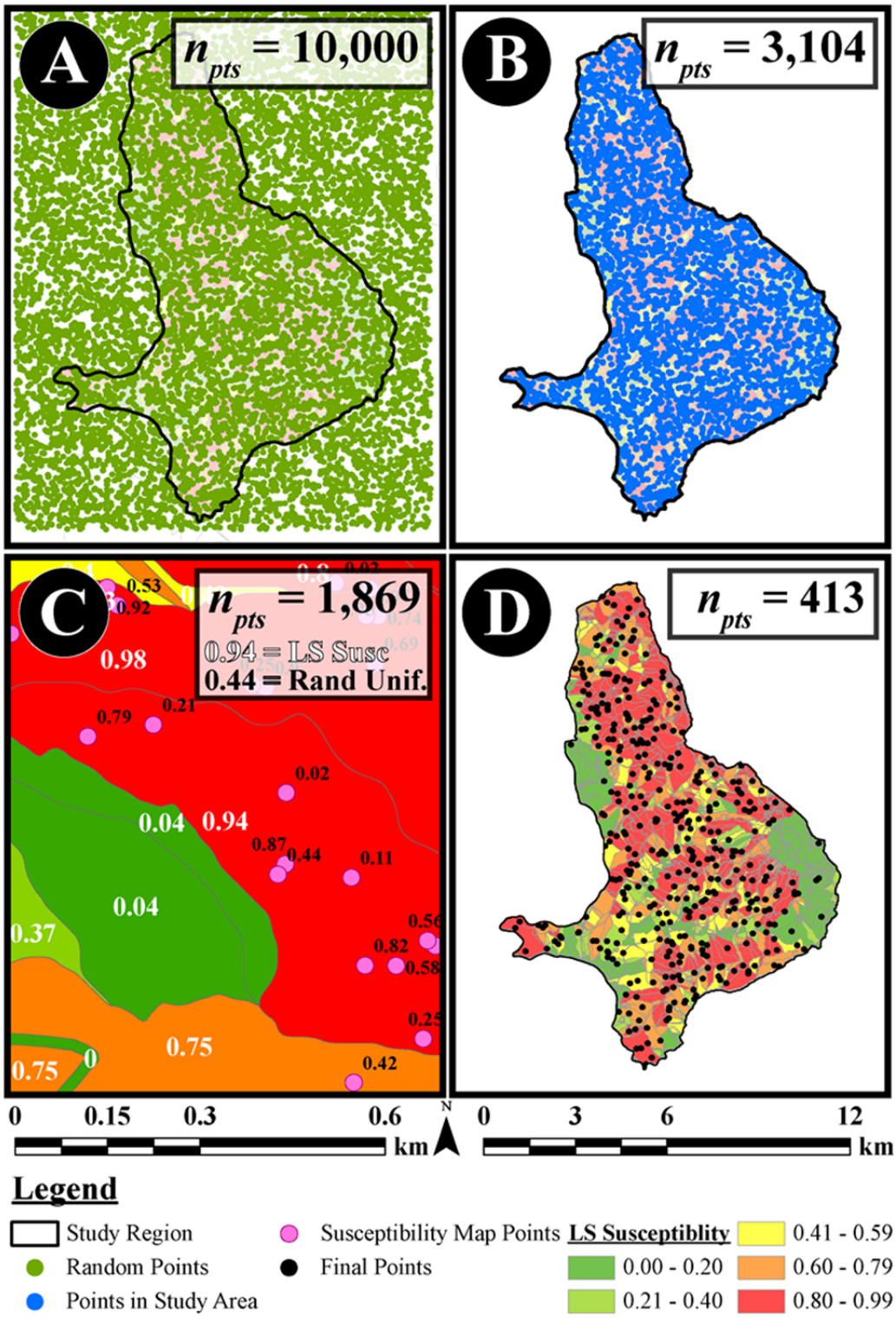


Figure 7.21 *Demonstration of how semi-randomly generated points are dropped for each model iteration in LRIM.* ***(A)*** *Random (N, E) locations are generated for* $n = 10 \times N_{LS}$*, where* N_{LS} *is the desired number of synthetic landslides. The n (N, E) coordinates are selected from uniform distributions across a rectangular extent encompassing the study region.* ***(B)*** *Any randomly generated points lying outside of the study region polygon are rejected.* ***(C)*** *For all remaining points, a random uniform value* $0 \leq x \leq 1$ *is generated and compared to the landslide susceptibility value* $0 \leq S_{LS} \leq 1$ *at this location. If* $x > S_{LS}$ *at this location, the point is rejected. Otherwise, the point is preserved. Due to the large number of points randomly generated, generally at this step there are more points than required as specified by the landslide density. Many of the points are randomly rejected to arrive at the desired number of landslide points as shown in* ***(D)****.*

At this stage of LRIM process, a map of $n = N_{LS}$ points has been generated and created. Each point has an associated value of landslide area (A_L) and length to width ratio (L/W) which will be used in **Method 7** and **8** to create a map of synthetic landslides simulating one iteration of an inventory of landslides resulting from a triggered landslide event.

7.3.6 Synthetic Landslide Orientation (Methods 7A and 7B)

From **Method 6**, a map of N_{LS} landslide (N, E) point locations has been created. In **Method 7A**, the value of slope aspect at that point location (created in **Method 2)** is queried and this value of aspect used as the orientation of the long axis of the landslide in **Method 7B** (**Figure 7.22**). This is to make the inventory more visually realistic, so that landslides appear to travel downslope (although we acknowledge that a small proportion of landslides in the real world are actually wider than they are long (Gabet and Dunne, 2002; Rickli *et al.*, 2008; Marchesini *et al.*, 2009; Milledge *et al.*, 2014)).

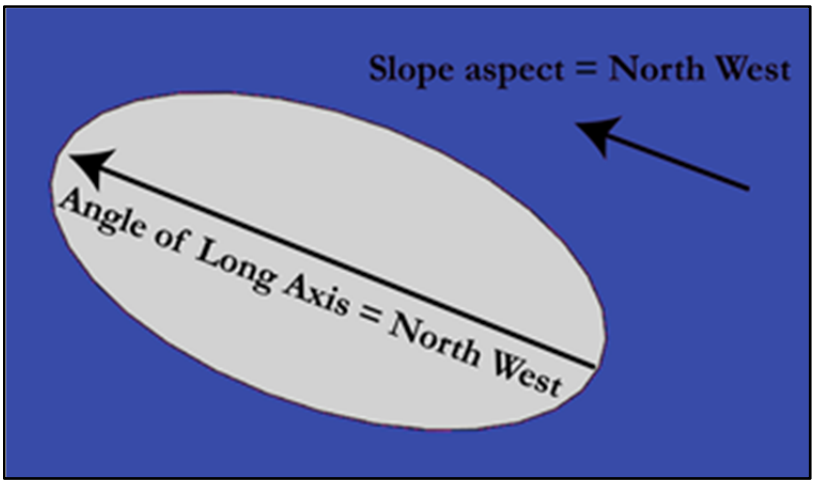


Figure 7.22 *Example of aligning long axis (length, L) of the landslide ellipse with average aspect of that slope unit.*

In **Method 7B**, each point location is buffered to create a landslide area using the GRASS GIS v.buffer tool (**Figure 7.23**). The buffer distances along the major and minor axes are calculated from landslide area (A_L) and length to width ratio (L/W) generated in **Methods 4** and **5**. The orientation of the major axis is defined in **Method 7A**. This process is repeated for each landslide individually.

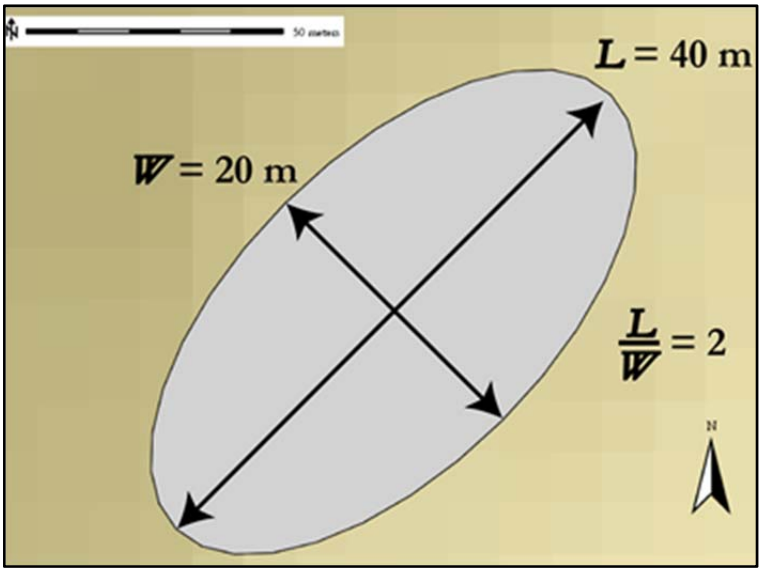


Figure 7.23 *Example of buffering a randomly generated landslide point location to create an elliptical polygon. Both the ellipse landslide area (A_L) and length to width ratio (L/W) are randomly selected from probability density functions, and from these the length (L) and width (W) of the ellipse calculated, with the ellipse plotted. The scale shown is 50 m.*

7.3.7 Synthetic Landslides Crossing Slopes and Creation of Synthetic Landslide Inventory Map (Method 8) (Output 1)

At this stage, LRIM has generated $n = N_{LS}$ individual landslide polygons and placed them on one synthetic inventory map. Due to the semi-random nature of how these landslides are 'dropped' over a study region, some landslides may cross ridges or valley bottoms (i.e., travel downhill and then uphill again), which is physically unrealistic for all but the very largest low mobility landslides (Crosta *et al.*, 2004) (shown in **Figure 7.24A**).

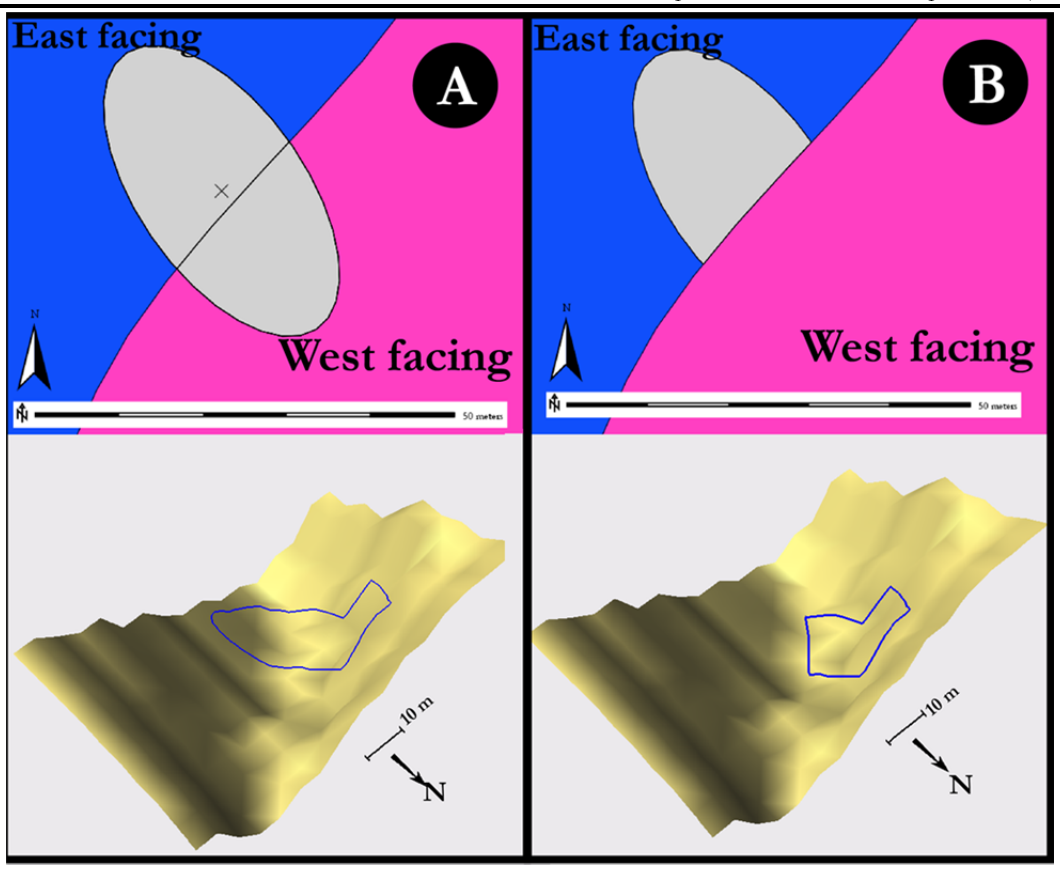


Figure 7.24 *Example of an LRIM generated synthetic landslide (not the final output of the model) crossing a slope unit division with a significant change in aspect between the two slope units.* ***(A)*** *The landslide point has been randomly generated (cross) and buffered to create a landslide area. This landslide crosses two slope units (pink and blue in A), which is unrealistic as the landslide travels downhill and then uphill (shown in the 3D visualisation).* ***(B)*** *The landslide has been "cut" so that only the portion of the landslide area remains that is within the original slope unit in which the point was dropped. The scale shown in the upper part of each figure A and B is 50 m, and in the lower part 10 m.*

In order to prevent landslides travelling down and then uphill (**Figure 7.24A**), each landslide polygon created in **Method 7B** is overlaid with a map of slope units, either input from the susceptibility map (**Input 2**) or created using a pre-existing Web Mapping Service (WMS), described in **Section 7.3.3.2**, **Method 2**. If the landslide polygon is completely contained by a slope unit (i.e., it does not cross a slope unit division), the LRIM continues on to the next stage. If a landslide polygon is intersected by two or more slope units, the landslide polygon is split into segments and the average azimuth of each slope unit is queried (corresponding to different segments of the landslide). Starting at the segment containing the centroid of the landslide and working outwards to contiguous segments, if the difference in azimuth between segments is $\leq 90°$, that segment is accepted. If the difference in azimuth is $> 90°$, it is deemed that it would be unrealistic for a landslide to travel across that slope unit, and that segment of the landslide is removed (**Figure 7.24B**). By working outwards from the centroid segment to contiguous segments, this ensures that the landslide remains in the original slope unit that it was dropped in, and that the landslide

is not split into separate pieces. For the Collazzone region (Umbria, Central Italy) (this region will be described in more detail in **Section 7.4.1**), this reduced the overall total landslide area by 5 % on average, so was found to not significantly alter the underlying probability density function of landslide areas.

The assumption that landslide run-out will cease when reaching a divide is unrealistic if modelling high mobility landslides, e.g., debris flows which tend to become channelized (Hungr *et al.*, 2005). In this case, we would expect the landslide to travel down slope and then turn into the channel, to roughly approximate an 'L' shape. In early stages of the LRIM development (before approximating landslides as ellipses described in **Chapter 5** and **Section 7.3.4.3**) Michele Santangelo and Ivan Marchesini (CNR-IRPI) introduced a modification to LRIM where landslide shape would flow through the least cost path in a digital elevation model (DEM), using the GRASS GIS tool 'r.drain' (Miller *et al.*, 2004). In this workflow, the landslide initiation point would be generated as per **Method 3** (**Section 7.3.4.1**), and then the DEM would be queried for the least cost line path downslope into the drainage channel. This line could then be buffered to create an irregularly shaped landslide area. Although this method may more realistically reflect the shape and flow behaviour of a debris-flow type landslide, we found that this was extremely computationally demanding, and could take several minutes to process per landslide, which was not practical for the large number of landslides simulated in LRIM. Potentially, LRIM could be developed to include this method in a more effective way (e.g., creating a map of least cost flow paths before the model is run).

In **Method 8B**, all individual landslide polygon maps are combined into one map to create a synthetic triggered landslide event inventory (**Output 1**). An example is shown in **Figure 7.25**.

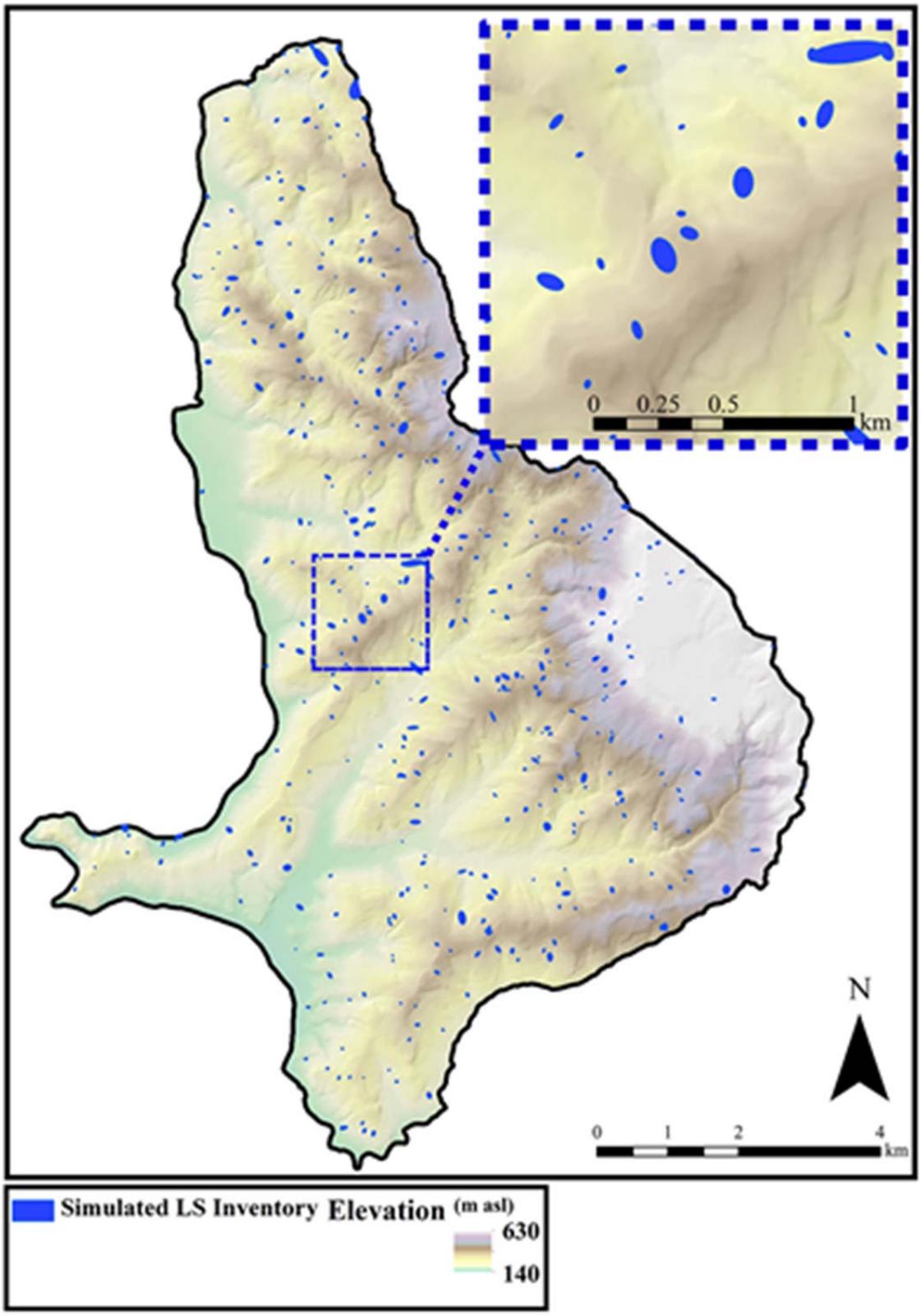


Figure 7.25 *One of hundreds of synthetic triggered landslide event inventory maps created by the Landslide-Road Impact Model (LRIM). In this example, N_{LS} = 413 landslides were dropped across the 79 km² Collazzone region (Umbria, Central Italy). The scale shown in the main figure is 4 km, and the inset figure 1 km.*

7.3.8 Overlaying Synthetic Triggered Landslide Event Inventory Maps with Roads (Method 9) (Outputs 2 and 3)

In **Method 9A**, the synthetic triggered landslide event inventory map (**Output 1**) created in **Method 8** is overlaid with a map of the road network (**Input 1**). Because the road map is a vector line, this does not accurately represent that roads are actually areas (i.e., they have a width). To take this into account, the synthetic triggered landslide event inventory is also overlaid with two buffered versions of the road map. The first is where all road lines have been buffered by 50 m either side of the line (total buffer zone 100 m) and the second is where all roads have been buffered by 100 m either side of the line (total buffer zone 200 m). This also takes into account that a landslide in close proximity to a road may cause disruption, even if it does not touch the road (Guzzetti *et al.*, 2003a), and small positional mismatches between spatial datasets. These buffer widths are based on typical distances in the literature reviewed in **Section 7.2** (e.g., Ardizzone *et al.*, 2014b; Brenning *et al.*, 2015).

Method 9B creates a map of lines corresponding to blocked segments of the road (**Output 2**) (**Figure 7.26**). **Method 9C** creates two maps of areas corresponding to the intersection between landslides and buffered roads (50 m and 100 m) (**Figure 7.27**) (**Output 3**).

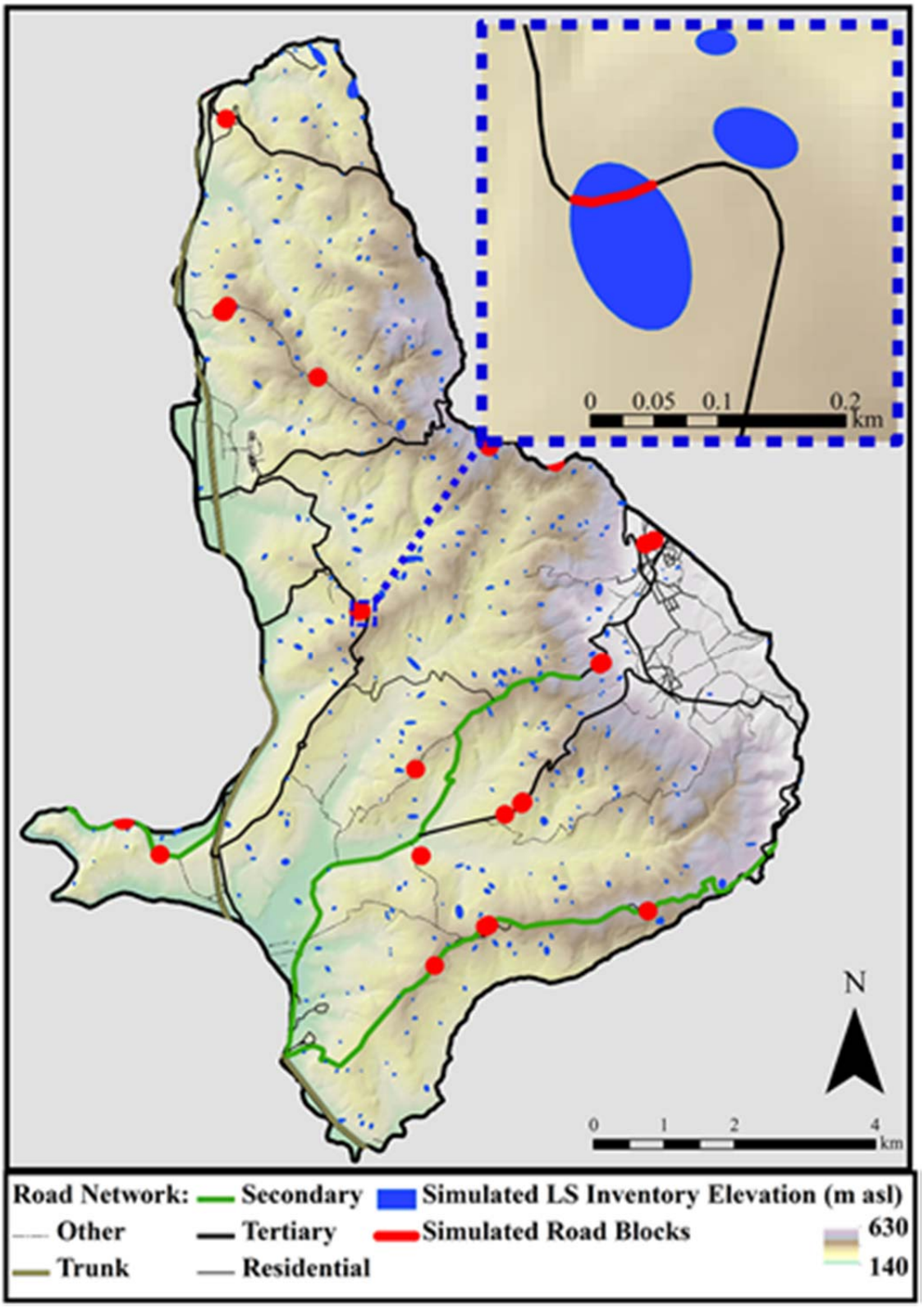


Figure 7.26 *One of hundreds of synthetic triggered landslide event inventories overlaid with the regional road network. Where a landslide polygon intersects with a road line, this is counted as a road blockage (marked in red). The scale shown in the main figure is 4 km, and the inset figure 0.2 km.*

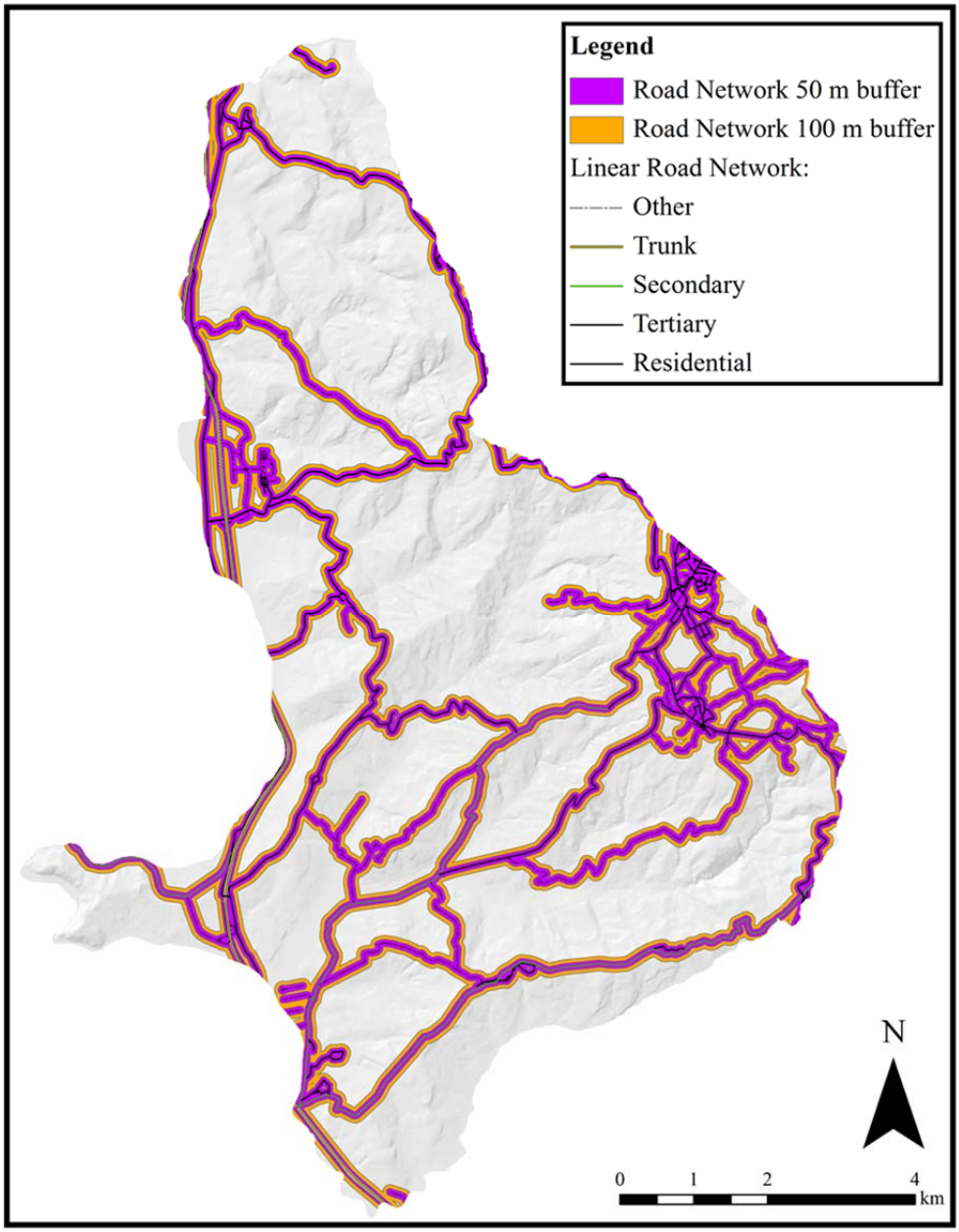


Figure 7.27 *Road network map (blue very thin lines) of Collazzone (Umbria, Central Italy) buffered by 50 m and 100 m (purple and orange areas) on each side of the roads. The vector line roads (thin blue lines) are taken to have no 'width' in the model. The scale in the figure is 4 km.*

In **Method 8**, the synthetic triggered landslide event inventory (**Output 1**) was created. In **Method 9**, maps of road blockages caused by landslides (**Output 2**) and landslides within 50 m and 100 m of the road (**Output 3**) were created. **Methods 3** to **9** are then repeated multiple times (we choose 100 iterations in our current version of the model) in a Monte Carlo type simulation and results are stored in a postgresql database (**Figure 7.28**).

The number of iterations was selected to be 100 as a trade-off between reasonable times to run LRIM and model robustness. By Robey and Barcikowski (1992)'s definition, 100 iterations of the model is not strongly robust for estimating the 'true' behaviour of the

system and the tails of the distributions of the outputs should be treated with some caution. Certainly, if LRIM is to be used for civil protection purposes, we would recommend $>10^3$ iterations of the model for more robust estimates of road network impact.

In each simulation, the landslide (*N, E*) locations, areas (A_L) and length to width ratios (L/W) will be slightly different, although still follow the underlying probability density functions described in **Methods 4** and **5**. After all iterations are complete, **Outputs 1** to **3** are aggregated and analysed to understand the overall statistics of the impact of a triggered landslide event scenario on the regional road network (**Output 4**).

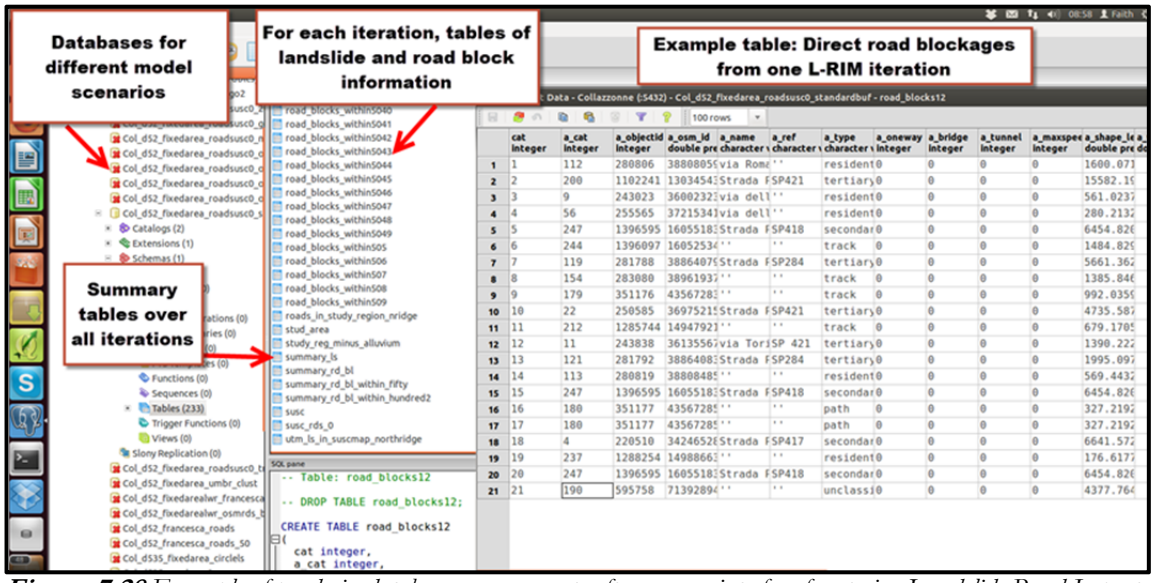


Figure 7.28 *Example of pgadmin database management software user interface for storing Landslide-Road Impact Model (LRIM) output. A new database is created for each simulation. Within this, numeric and spatial information about the landslides and road blockages (**Outputs 1** to **3**) are stored. This allows relatively simple storage, management and analysis of very large datasets.*

7.4 Study Regions

Here detailed descriptions are given for three study areas for where the Landslide-Road Impact Model (LRIM) has been applied: the Collazzone region (Umbria, Central Italy) (**Section 7.4.1**); the Oat Mountain Quadrangle (Northridge, California, USA) (**Section 7.4.2**) and the Su-Hua region (Northeast Taiwan) (**Section 7.4.3**). For each description, we also include reference to the LRIM **Inputs 1** to **6** described in **Section 7.3**.

Due to the 'generally applicable' statistical patterns of landslide size and shape used in LRIM (**Methods 4** and **5**), LRIM can theoretically be used to simulate the impact of triggered landslide events on road networks in any medium-high topographic region subject to low mobility landslides. LRIM was developed using two study areas: Collazzone

region (Umbria, Central Italy) (**Section 7.4.1**) and Oat Mountain (Northridge, California, USA) (**Section 7.4.2**) and then applied to Su-Hua (Northeast Taiwan) (**Section 7.4.3**). These regions were chosen for the following reasons:

- *Collazzone* (Umbria, Central Italy) is a site where **Inputs 1** to **5** are available and there is local expertise to inform **Input 6**.
- *Oat Mountain* (Northridge, USA) is a site where **Inputs 1** to **5** are available and characteristics are different to Collazzone (in terms of topography, road network density).
- *Su-Hua* (Northeast Taiwan) is a site where **Inputs 1, 4** and **5** are available and characteristics are different to Collazzone and Northridge in terms of topography and economic status of the country (assuming this has an impact on the approach to road building). This presents an opportunity to apply LRIM to a 'low data' scenario.
- *All three sites* have experienced triggered landslide events in recent decades and triggered event landslide inventory maps are available. These triggered landslide events are simulated in LRIM (by setting landslide density to match observations) and the real triggered event inventory maps are used to confront LRIM output data. The triggered events that we use to confront our model with, have three different types of triggering mechanisms (Collazzone—snowmelt; Northridge—an earthquake, Su-Hua—a typhoon).

The sources and information about LRIM **Inputs 1** to **6** and triggered event inventory data for the Collazzone (Umbria, Central Italy), Oat Mountain (Northridge, USA) and Su-Hua (Northeast Taiwan) sites are presented in the following sections.

7.4.1 Collazzone, Central Italy

The Collazzone study region is a 79 km^2 drainage basin, irregularly shaped, in the Umbria Region of Central Italy, located in the Province of Perugia (**Figure 7.29**). It extends from about 42.8° to 42.9 *N* and 12.4° to 12.5° *E*, over an area approximately 16.34 km (*N*–*S*) × 11.39 km (*E*–*W*). The Collazzone study region follows the Tiber River basin, from the flood plain to Monte di Grutti (East to West) and Tiber river tributaries to the North and South (Guzzetti *et al.*, 2006).

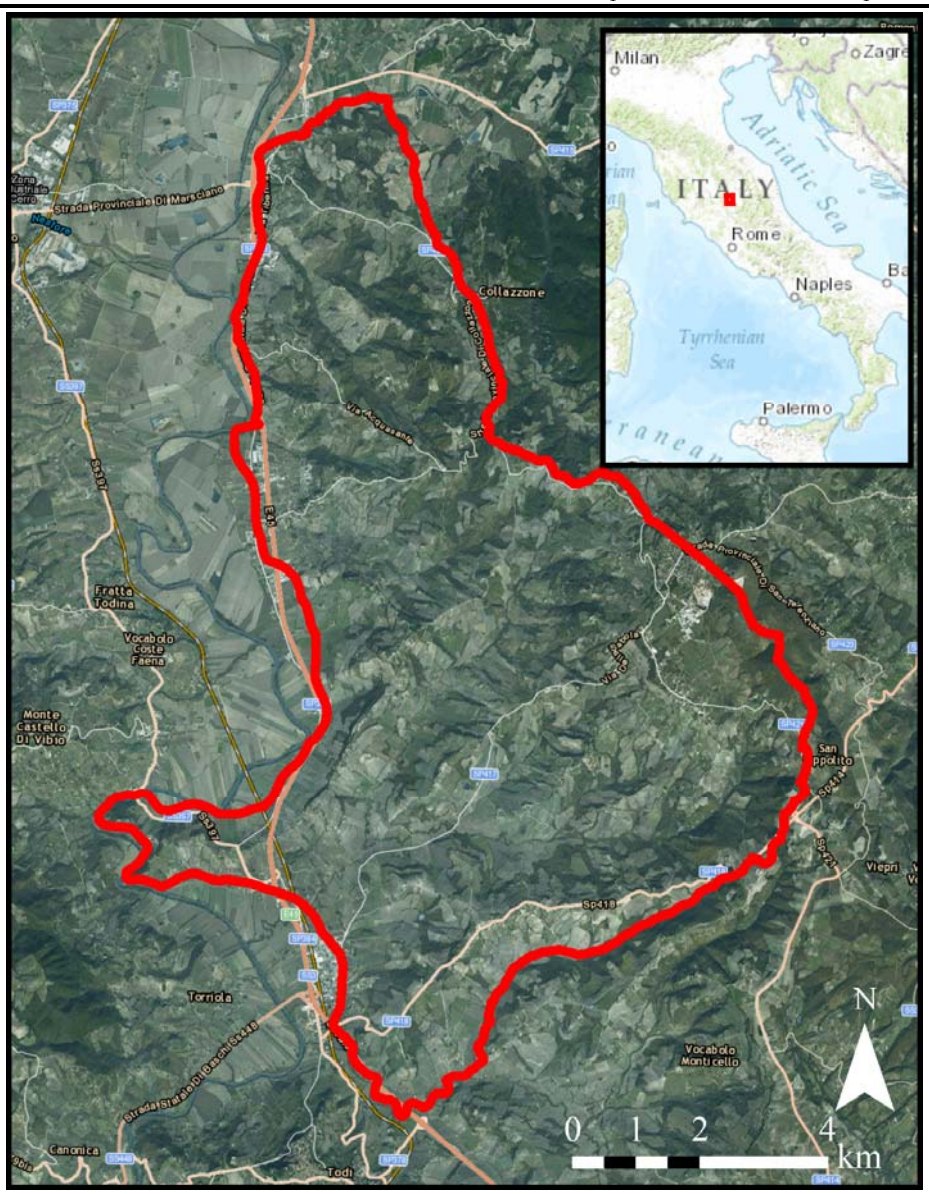


Figure 7.29 *Overview map of the Collazzone (Central Italy) study region. The red outline shows the extent of the study region. Base maps (imagery, roads and gazetteer) from ESRI et al. (2013). Study region extent from Guzzetti et al. (2006a).*

7.4.1.1 Road Network Map (Input 1)

Various road maps of varying detail and quality were available, shown in previously **Figure 7.12** (**Section 7.3.2.1**). Two versions of the road network were used in the model:

- *Open street map (OSM, 2014) (**Figure 7.12C**)*. The 153 km vector line map of roads created by Open Street Map is a collaborative project to create open source, freely available maps for all countries across the globe. Features are mapped by a team of >33,000 registered volunteers from satellite imagery, GPS tracks combined with free sources of spatial data (Haklay and Weber, 2008). In regions of the world with a large core of volunteer mappers, OSM provides some of the most detailed, up-to-date maps that are freely available. Roads are classified by type, from unpaved track through to highway, although in some

cases, data may be inaccurate depending on the mapping method. Coverage can be inhomogeneous if areas have not been mapped, or combined data sources are used. Road widths are not included but can be estimated from typical European standards (e.g., Murphy, 2011).

- *Vector polygons of road area mapped from aerial photography at the time of a triggered landslide event* (Ardizzone *et al.*, 2008) (**Figure 7.12E**). The 400 km vector polygon map of roads has the benefit of being expertly mapped to a very fine level of detail (e.g., including temporary farm tracks) with consistent coverage over the Collazzone basin, although roads are not separated by hierarchy. However, this map is not freely available to download, and it is unlikely such detailed maps exist for other regions where the LRIM may be applied. This map gave more detailed insight into the interaction between landslides and roads during LRIM development.

7.4.1.2 Landslide Susceptibility Map (Input 2)

A susceptibility map (shown previously in **Section 7.3.2**, **Figure 7.11B**) from Rossi *et al.* (2010) was used. Their map of shallow landslide susceptibility for the Collazzone Basin, Central Italy was produced at the slope unit scale through linear discriminant analysis of 51 variables including terrain and land cover, and they then compared this to a multi temporal inventory of 2,455 landslides that occurred between 1941 and 1996 as a result of various triggers. They then validated this against a 9-year multi temporal inventory of landslides from 1997 – 2005. Rossi *et al.* (2010a) created a polygon vector map of slope units by constructing a map of drainage and divide lines from the 10 m × 10 m DEM, using the method of Carrara *et al.* (1991; 1995). There are 894 slope units in the region, of the order of 100 m^2 to 1 km^2. The median slope unit area is 50,000 m^2.

7.4.1.3 Study Area Boundary Map (Input 3)

The source of the study area boundary map (shown previously in **Section 7.3.2**, **Figure 7.11**) was Guzzetti *et al.* (2006a).

7.4.1.4 Digital Elevation Model (Input 4)

We used a DEM (Shown previously in **Section 7.3.2**, **Figure 7.11C**) described by Ardizzone *et al.* (2007), where the 10 m DEM was created using a 1:10,000 scale topographic map of the area, interpolating 5 m and 10 m contour lines.

7.4.1.5 Landslide Density (Input 5)

Landslide density (L_D) was set to match a real triggered landslide event inventory, described below in **Section 7.4.1.7**. This was 5.13 LS km^{-2}. Other scenarios of landslide density have also been explored in LRIM (results shown later, at the end of **Section 7.5.1**).

7.4.1.6 Broad Characteristics of Road Landslide Protection in Region (Input 6)

Using the road network maps (**Input 1**) and an existing triggered event inventory (**Section 7.4.1.7**, below), we investigated the distances between roads and landslides for different hierarchies of road, discussed in **Section 7.3.2.6**. This data was used to establish what proportion of the road network length to randomly select to buffer and what widths those buffers should be for LRIM **Input 6** when running in **Mode C** (reduced landslide susceptibility near roads). Values derived for Collazzone are presented in **Table 7.5**.

Table 7.5 *Road buffer widths and proportion of the road network to randomly select to buffer for Collazzone, Input 6. Buffer widths are based on second percentile of all observed distances (D) between observed landslide centroids and the nearest road type for the Collazzone region (**Section 7.3.2.6**). % of road network to be buffered is based on the percentage of the landslide inventory nearest to that road type. Methodology described in **Section 7.3.2.6**.*

Road Hierarchy	Buffer width (m)	% of Road Network Buffered
Primary	22	98
Secondary	63	80
Tertiary	41	60
Minor	5	60

7.4.1.7 Triggered Landslide Event Inventory (Model Confrontation Data)

A triggered landslide event inventory from the Umbria region (Central Italy) is used to confront the LRIM model output. In this triggered landslide event (Cardinali *et al.*, 2000) there was rapid snowmelt on the 1 January 1997, which triggered thousands of landslides across an area of about 2,000 km^2 in the Umbria region—within which lies the Collazzone region. Cardinali *et al.* (2000) mapped 4,233 landslides by using extensive field reconnaissance in the weeks after the event and aerial photography flown three months after the event, at a nominal scale of 1:120,000. About 60 % of landslides were shallow slips and slumps, and 38 % deep seated. Our model confrontation data is a subset of this larger event inventory, consisting of 413 landslides in the Collazzone region. Observed landslide density in the Collazzone basin was 5.13 LS km^{-2} for this event. A figure for this inventory is presented later in **Section 7.5.2** (Model Confrontation).

7.4.2 Oat Mountain Quadrangle, Northridge, USA

The Oat Mountain 7.5' Quadrangle, Northridge, in California USA extends over an area of approximately 13.90 km (*N* to *S*) × 11.49 km (*W* to *E*) (34.25° to 34.38° N and –118.63° to –118.50° W) (**Figure 7.30**).

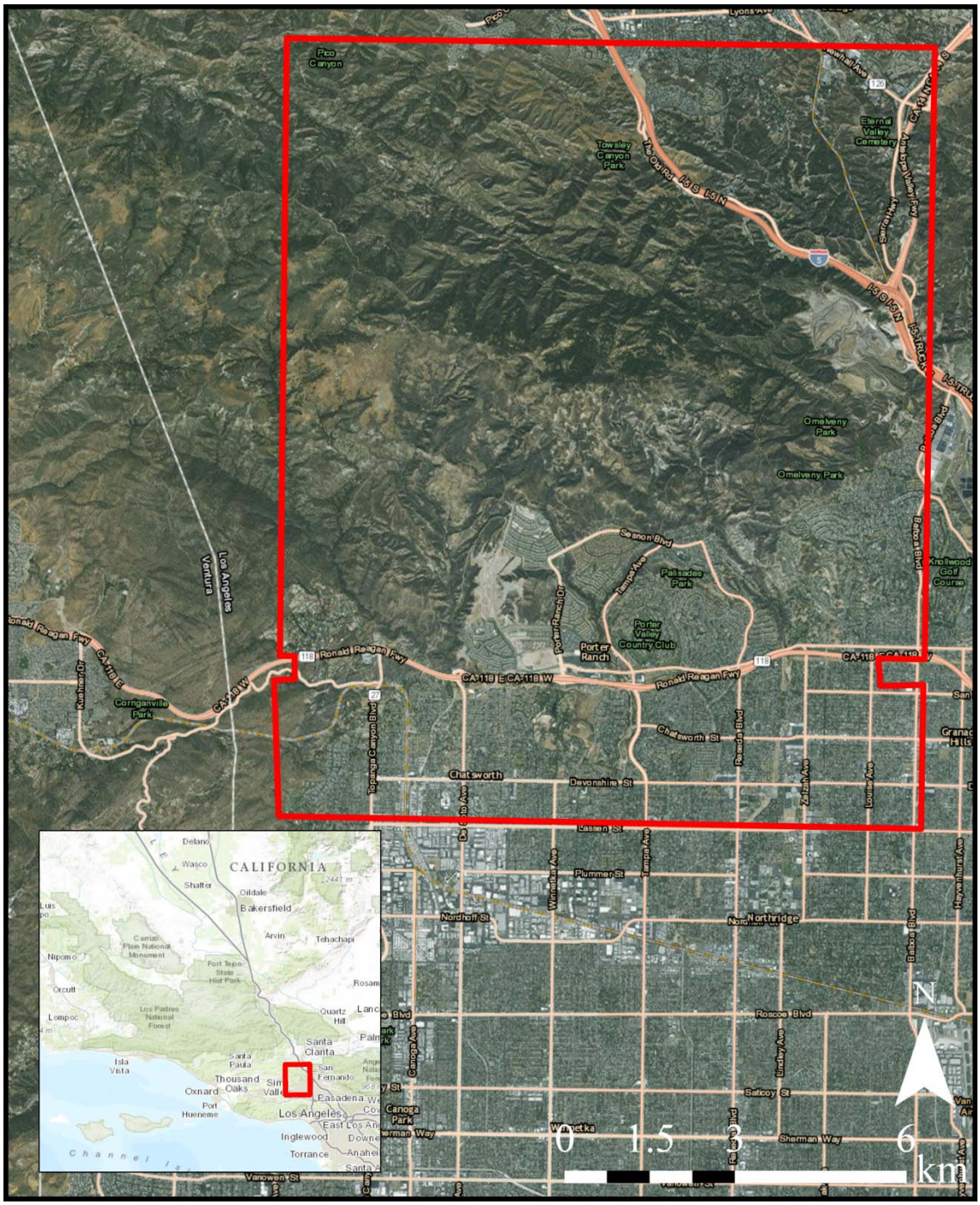


Figure 7.30 *Overview map of the Oat Mountain (Northridge, USA) study region. The red outline shows the extent of the study region. Base data (imagery, road map and gazetteer) from ESRI et al. (2013).*

7.4.2.1 Road Network Map (Input 1)

This road network map (**Figure 7.31A**) was obtained from the United States Census Bureau (USCB, 2014) and consists of a 780 km vector line file of major and minor roads; it was last updated in 2014. Homogeneous maps of roads, other infrastructure and places are available for the entire USA and freely available to download. Roads are classified by hierarchy, and range from unpaved tracks to highways. Road widths are not included but can be estimated from typical road widths (FHWA, 2011).

7.4.2.2 Landslide Susceptibility Map (Input 2)

A landslide susceptibility map (**Figure 7.31B**) for this region was obtained from Jibson *et al.* (1998) where a 10 m × 10 m spatial resolution raster susceptibility map was based on an infinite slope model and inventory of landslides triggered by the Northridge Earthquake. Susceptibility is split into 7 classes. This map is available as a non-georeferenced tiff file, georeferenced manually from visual identification of ground control points – there were some small errors in spatial alignment. In addition, a vector polygon map of slope units (**Figure 7.31C**) was created using a 30 m DEM (**Input 4**) and a map of alluvial units (USGS, 2005) input into a web processing tool developed by Alvioli *et al.* (2014). Parameter values were based on experimentation from default values. Final values used were: 800 (30 m × 30 m) cells for basin creation, reduction factor = 9 (the factor by which the slope unit divisions are reduced by in each iteration of the tool until the slope units match user defined specifications), maximum standard deviation of sin/cos aspect per unit = 0.5 rad, minimum half basin area = 0.80 km^2.

7.4.2.3 Study Area Boundary (Input 3)

The study area boundary (**Figure 7.31C**) was obtained from Jibson *et al.* (1998) where a 155 km^2 vector polygon map of study region was created from the extent of the landslide susceptibility map. Due to issues with the georeferencing process, two small regions of the susceptibility map extent were removed from the study area boundary, visible in the lower portion of **Figure 7.31C**.

7.4.2.4 Digital Elevation Model (Input 4)

The source of the DEM (**Figure 7.31C**) was USGS (2006), who created this 30 m × 30 m raster of elevation from satellite radar imagery of the Earth's surface, from the Shuttle Radar Topography Mission (SRTM).

7.4.2.5 Landslide Density (Input 5)

Landslide density (L_D) was set to match a real triggered landslide event inventory, which is described below in **Section 7.4.2.7**. This was 8.70 *LS* km^{-2}.

7.4.2.6 Broad Characteristics of Road Landslide Protection in Region (Input 6)

Using the road network maps (**Input 1**) and an existing triggered event inventory (**Section 7.4.2.7**, below), we investigated the distances between roads and landslides for different hierarchies of road, discussed in **Section 7.3.2.6**. This data was used to establish what proportion of the road network length to randomly select to buffer and what widths those buffers should be for LRIM **Input 6** when running in **Mode C** (reduced landslide susceptibility near roads). Values derived for Oat Mountain are presented in **Table 7.6**.

Table 7.6 *Road buffer widths and proportion of the road network to randomly select to buffer for Collazzone,* ***Input 6****. Buffer widths are based on the second percentile of all observed distances (D) between observed landslide centroids and the nearest road type for the Northridge inventory (****Section 7.3.2.6)****. % of road network to be buffered is based on the percentage of the landslide inventory nearest to that road type. Methodology described in* ***Section 7.3.2.6****.*

Road Hierarchy	Buffer width (m)	% of Road Network Buffered
Primary	26	99
Secondary	8	97
Tertiary	10	14
Minor	22	88

7.4.2.7 Triggered Landslide Event Inventory (Model Confrontation Data)

A triggered landslide event inventory from Northridge, California (USA) is used to confront LRIM model output. In this triggered landslide event (Harp and Jibson, 1995) there was an earthquake on 18 January 1994 (M_W = 6.7), which triggered thousands of landslides across an area of about 10,000 km^2—within lies the Oat Mountain Quadrangle. Harp and Jibson (1995) mapped 11,111 landslides immediately after the earthquake using field surveys and 1:60,000 scale aerial photography. Over 90 % of the mapped landslides were small shallow failures A vector polygon map of 1,356 landslides in the Oat Mountain Quadrangle region is used. Observed landslide density in the Oat Mountain Quadrangle was 8.70 *LS* km^{-2} for this event. A figure for this inventory is presented later in **Section 7.5.2** (Model Confrontation).

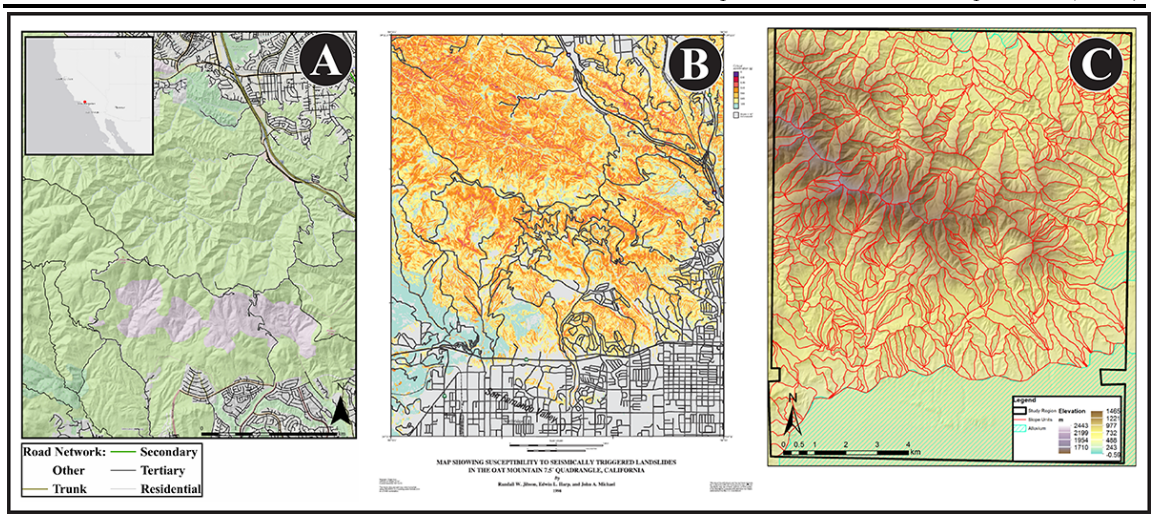


Figure 7.31 ***Inputs 1*** *to* ***4*** *for the Oat Mountain Quadrangle (Northridge, USA) Landslide-Road Impact Model (LRIM) site.* ***(A)*** *Road network map (USCB, 2014) (****Input 1****)* ***(B)*** *Landslide susceptibility map (Jibson et al., 1998) (****Input 2****).* ***(C)*** *Digital Elevation Model (USGS, 2006) (****Input 4****) overlaid with map of slope unit divisions created for the Oat Mountain Quadrangle using a web processing service tool developed by Alivioli et al. (2014). The study area boundary (****Input 3****) is visible in (C). The scale in A and C are 4 km, and in B the lower scale is 2 km. Regions shown are slightly different in extent.*

7.4.3 Su-Hua, Northeast Taiwan

The Su-Hua study region is a 421 km^2 irregularly shaped zone, in the Northeast of Taiwan, approximately following the route of Provincial Highway 9 (**Figure 7.32**). The extent of the study region was defined manually based on the extent of a triggered landslide event inventory, described in **Section 7.4.3.7**. The region extends from about 24.2° to 24.6 *N* and 121.6° to 121.9° *E*, over an area approximately 43 km (*N*–*S*) × 24 km (*E*–*W*). The Su-Hua region represents a contrast to Collazzone (Italy) and Northridge (USA) as:

- The triggered landslide event inventory was caused by a Typhoon (described later in **Section 7.4.3.7**)
- The region is classified as 'upper-middle income' (World Bank, 2014) so may have different approaches to road building and consequently, landslide-road interactions (as discussed in **Section 7.2.2.4**)
- There is less 'high confidence' input data freely available for this region, representing an opportunity to test LRIM in a low-data scenario.

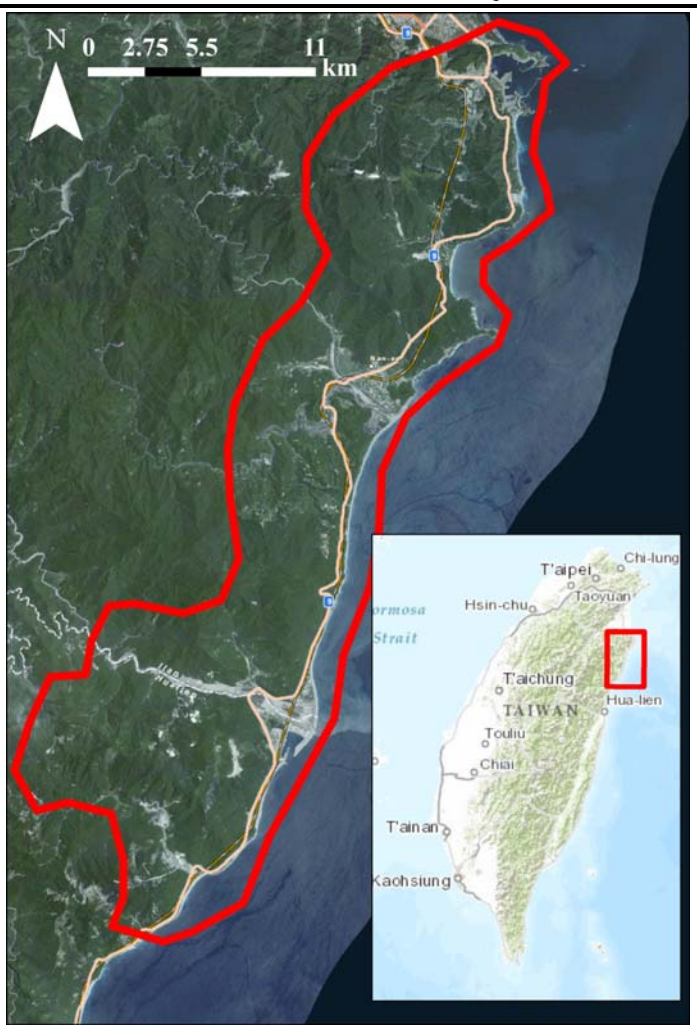


Figure 7.32 *Overview map of the Su-Hua (Northeast Taiwan) study region. The red outline shows the extent of the study region. Base data (imagery, road map and gazetteer) from ESRI et al. (2013).*

7.4.3.1 Road Network map (Input 1)

This road network map (**Figure 7.23A**) was obtained from Open Street Map (OSM, 2014) and consists of a 265 km vector line file of major and minor roads (more fully described in **Section 7.4.1.1**). Roads are classified by type, from unpaved track through to highway, although in some cases, data may be inaccurate depending on the mapping method. A visual assessment overlaying the OSM data in Google Earth suggests that most of the main roads are mapped in OSM, although some smaller roads leading to small villages may be missed. The road map is shown in **Figure 7.34A**.

7.4.3.2 Landslide Susceptibility Map (Input 2)

At the time of performing this research, a landslide susceptibility map was not available to us for the Su-Hua region. We attempted to create an approximate map of landslide susceptibility using SHALSTAB which is a deterministic model developed by Montgomery and Dietrich (1994) based on soil and slope parameters. Creation of the susceptibility map

was performed using the Grass-GIS module (r.shalstab) (Filipello and Strigaro, 2014), where the user inputs a digital elevation model (DEM) and soil parameters (listed in **Table 7.7**). The output is a landslide susceptibility map with seven discrete classes of probability of failure ranging 1 (unconditionally unstable) to 7 (unconditionally stable). A 5 m × 5 m DEM and suggested soil parameters for the region were provided by Yichin Chen from the Taiwan Typhoon and Flood Research Institute, National Applied Research Laboratories (personal communication, 2014), described in **Table 7.7**. These were used as inputs to r.shalstab. The output map of susceptibility was then divided by 7 (so that susceptibility ranges $0.0 \leq S_{LS} \leq 1.0$), and is shown in **Figure 7.33**. Note, other than visual analysis to confirm that the inventory of landslides (described later in **Section 7.4.3.7**) appeared to occur primarily in high susceptibility regions, the robustness of this susceptibility map has not been tested. Thus, it should only be used as an approximate guide as to where landslides occur throughout the region.

Table 7.7 *Typical soil parameters for the Shu-Wa region of Taiwan used as inputs to SHALSTAB (Montgomery and Dietrich, 1994) to calculate a map of landslide susceptibility.*

Parameter	Value or Input	Source
Elevation (m)	5 m × 5 m DEM Raster Map	(Chen, personal communication, 2014)
Soil Friction Angle, ϕ (°)	20.76	(Chen, personal communication, 2014)
Soil Cohesion, c (N m^{-2})	2,950	(Chen, personal communication, 2014)
Soil Density, g (km m^{-3})	22	(Chen, personal communication, 2014)
Vertical Thickness of Soil, z (m)	25.601 × exp(-0.096 × Slope Angle)	(Chen, personal communication, 2014)
Hydraulic Conductivity, k (m hr^{-1})	0.016	Estimated from typical soil type for the region given by (Lo *et al.* 2014) using tables from (Bear, 1972).

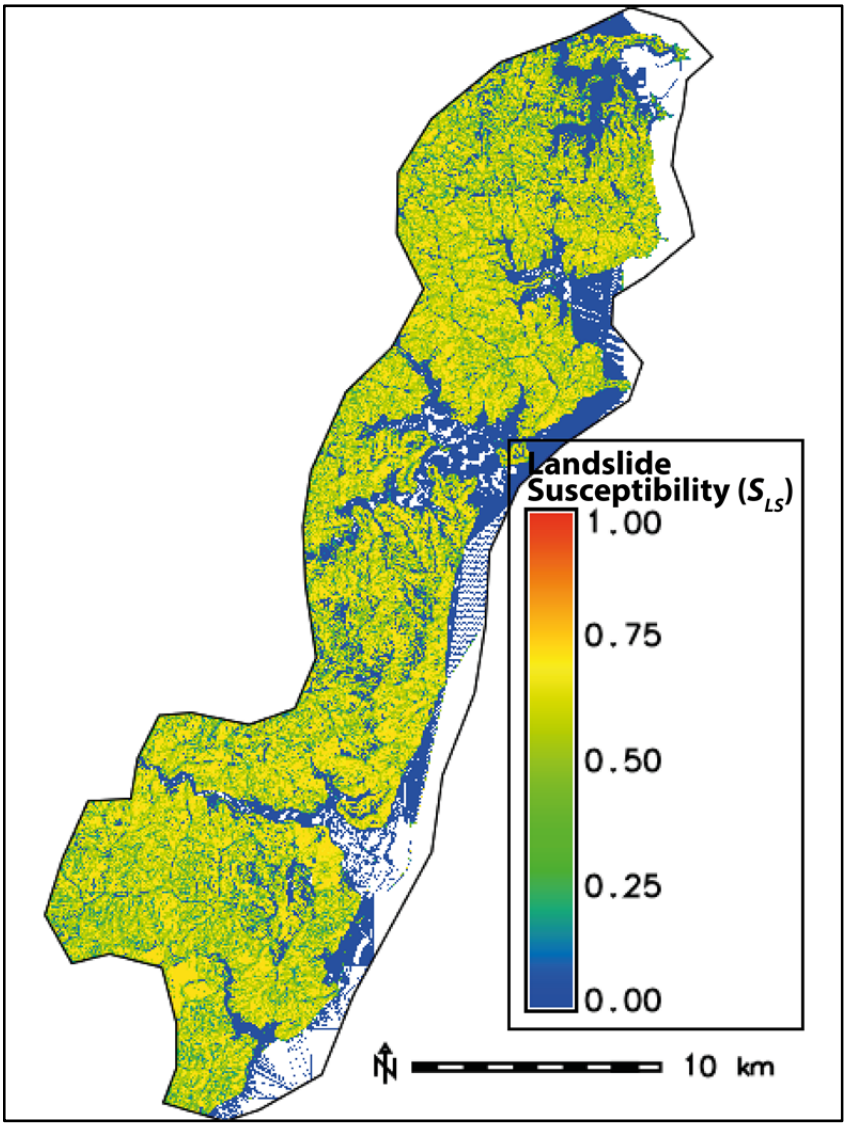


Figure 7.33 *Landslide susceptibility map for the Su-Hua study region created using SHALSTAB (Montgomery and Dietrich, 1994) using the Grass-GIS add on tool r.shalstab using parameter values listed in* ***Table 7.7.***

7.4.3.3 Study Area Boundary (Input 3)

The study area boundary was manually drawn to outline the extent of the triggered landslide event inventory we have available (**Section 7.4.3.7**). This is a 421 km^2 irregularly shaped polygon, shown in **Figure 7.34A, B, C**.

7.4.3.4 Digital Elevation Model (Input 4)

The source of the 5 m ×5 m digital elevation model was provided by Chen (personal communication, 2014). The DEM is shown in **Figure 7.34B**. A map of slope unit divisions was created using the same process and parameters as described for Oat Mountain (**Section 7.4.2.4**).

7.4.3.5 Landslide Density (Input 5)

Landslide density (L_D) was set to match a real triggered landslide event inventory, which is described below in **Section 7.4.3.7**. This was 0.35 LS km^{-2}.

7.4.3.6 Broad Characteristics of Road Landslide Protection in the Region (Input 6)

Input 6 was not investigated for the Su-Hua study region. We believe landslide occurrence may be higher in close proximity to roads for this site, and this is something we would like to investigate in detail at a later date.

7.4.3.7 Triggered Landslide Event Inventory (Model Confrontation Data)

A triggered landslide event inventory from Su-Hua, Northeast Taiwan is used to confront LRIM model output. In this triggered landslide event in October 2010, Typhoon Megi passed through the South China Sea at the same time as the Asian Monsoon, resulting in record high rainfall in Taiwan (Lo *et al.*, 2014), triggering landslides across the region. We have an inventory of 149 landslides mapped by Chen (personal communication, 2014) along the Su-Hua highway in the northeast of Taiwan. Observed landslide density across the Su-Hua region was 0.35 LS km^{-2} for this event. This inventory is shown in **Figure 7.34C**.

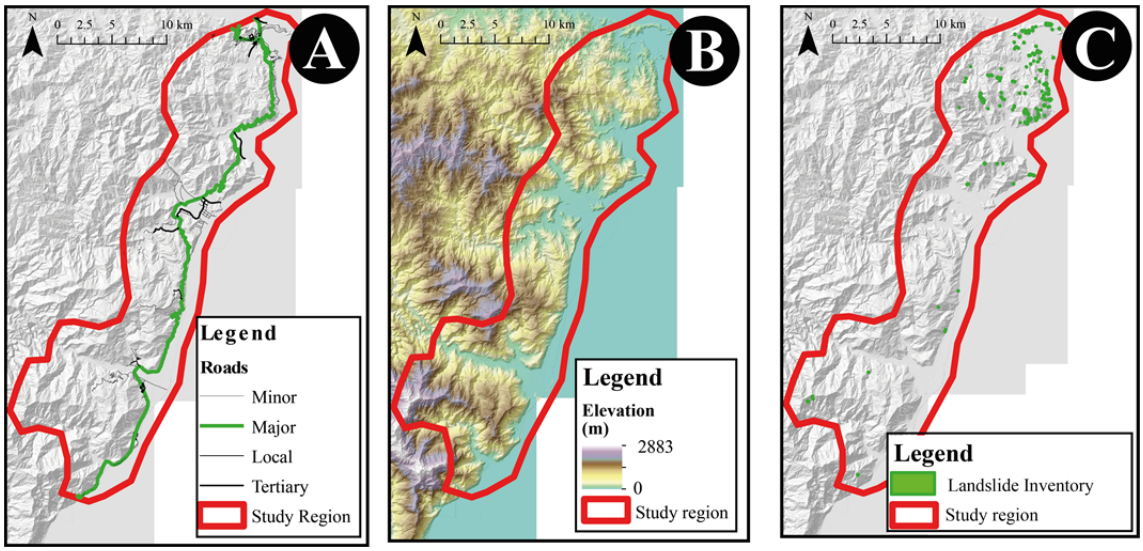


Figure 7.34 *LRIM* ***Inputs 1, 3 and 4*** *and model confrontation data for the Su-Hua region, Northeast Taiwan.* ***(A)*** *Road map of major and minor roads from Open Street Map (OSM, 2014) (****Input 1****) and study area boundary (****Input 3****)* ***(B)*** *5 m × 5 m digital elevation model (Chen, personal communication, 2014) (****Input 4****)* ***(C)*** *Inventory of N_{LS} = 149 landslides triggered by Typhoon Megi in October 2010 (Chen, personal communication, 2014).*

7.5 Results & Model Confrontation

In this section, examples of **Outputs 1** to **4** (**Section 7.5.1**) are given, followed by a summary the results of LRIM confrontation using observed data (**Section 7.5.2**). In

Section 7.5.3, other parameters varied in the model are discussed, followed by a preliminary discussion of road network analysis (**Section 7.5.4**).

7.5.1 Examples of Outputs 1 to 4

Figure 7.35 shows an example of a synthetic triggered landslide event inventory (**Output 1**) and resulting road blockages caused by landslides (**Output 2**) for one out of 100 LRIM iterations for Collazzone (Umbria, Central Italy) where the landslide density (L_D) = 5.13 *LS* km^{-2} (**Input 5**) and Road Landslide Protection set to 'standard' (i.e., no adjustment for landslide-road interactions) (**Input 6**). In this iteration, there were 52 direct road blockages (N_{BI}), 76 landslides within 50 m of the road and 126 landslides within 100 m of the road. Each iteration of LRIM (out of 100) will produce an output map of the location and size of road blockages across the region. This allows potential users to explore different scenarios of landslide-road network disruption resulting from the same 'magnitude' triggered landslide event (i.e., the total number of landslides is the same for each iteration but if several landslides occur on the major highway as opposed to minor roads, the disruption may be considerably higher).

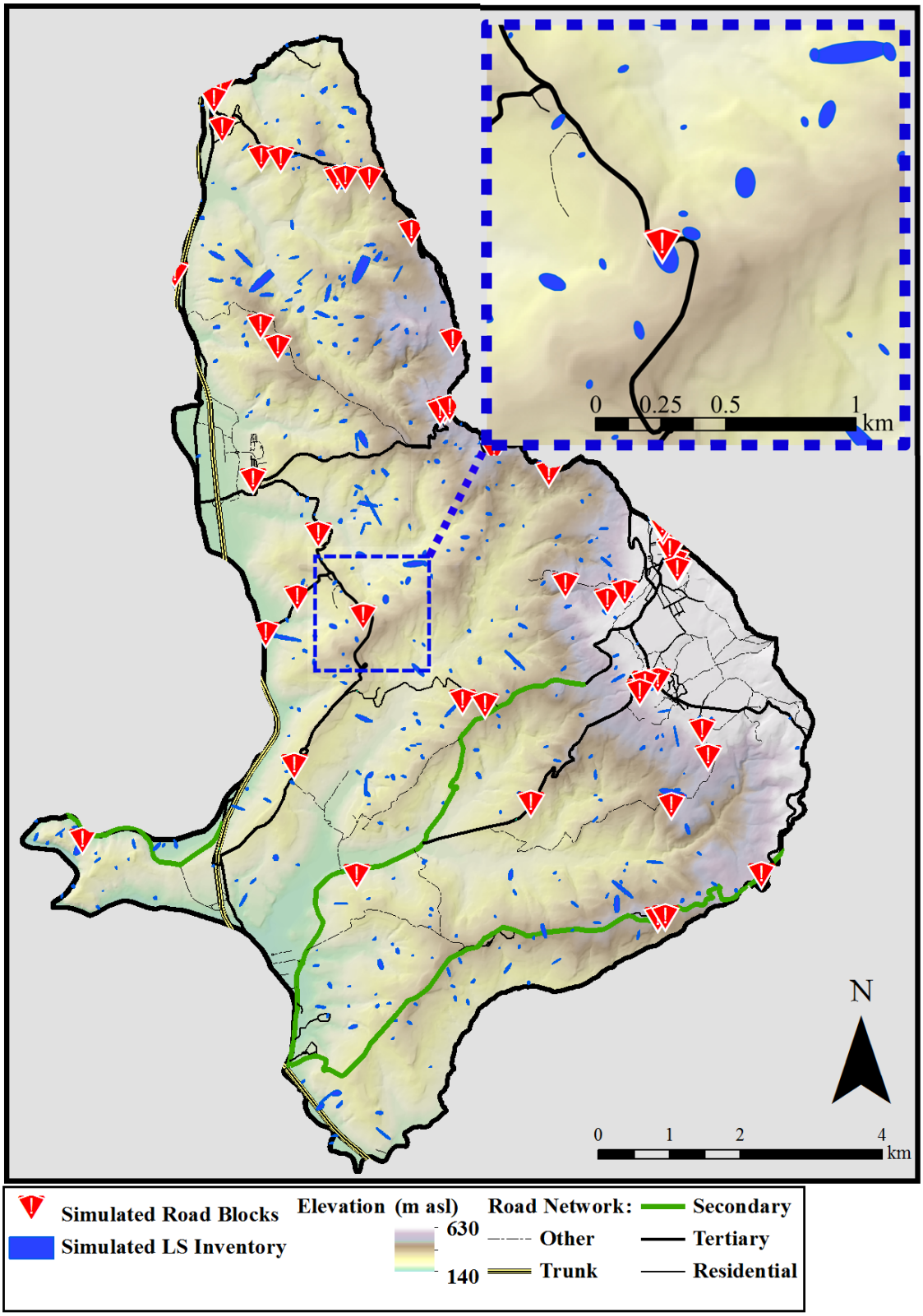


Figure 7.35 *Example Landslide-Road Impact Model (LRIM) output for the Collazzone region (Umbria, Central Italy) showing a simulated landslide inventory (blue) and resulting road blockages across major and minor roads. This map shows results from 1 iteration of 100. In this simulation there are 413 landslides resulting in 52 road blockages. The scale shown in the main figure is 4 km, and the inset figure 1 km.*

The LRIM output maps can also be converted to *.kmz files and draped over imagery in Google Earth. An example of a simulated road blockage on a minor road is shown in

Figure 7.36 by overlaying **Outputs 1** and **2** in Google Earth. Draping the LRIM output maps over 3D Google Earth imagery may be useful for visualising LRIM output in combination with other layers of spatial information — particularly where local thematic data (e.g. place names, road maps) is difficult to obtain or out-of-date.

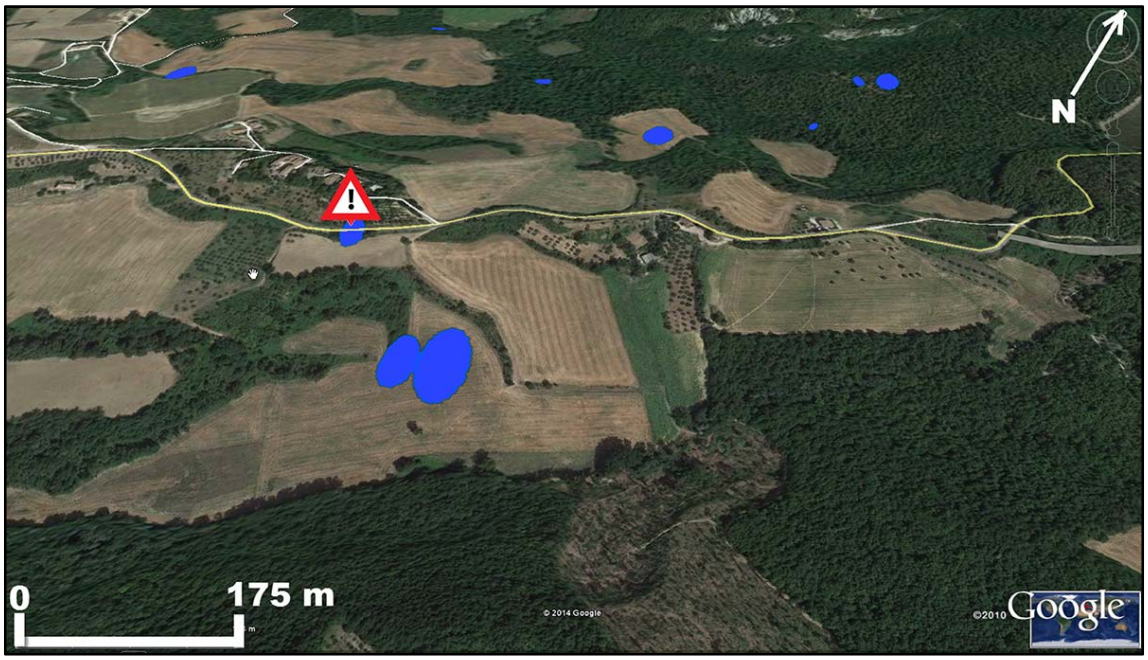


Figure 7.36 *Zoomed region of the Collazzone region site (Umbria, Central Italy) showing simulated landslides (blue) and resulting road blockages (red) (as in* ***Figure 7.35****) overlaid on Google Earth imagery. Image: Google Earth (2014).*

Figure 7.37 shows an example of a synthetic triggered landslide event inventory (**Output 1**) and the resulting direct road blockages caused by landslides (**Output 2**) for one out of one hundred iterations of LRIM for Oat Mountain (Northridge, USA) where the landslide density (L_D) = 8.70 LS km^{-2} (**Input 5**) and Road Landslide Protection set to 'standard' (i.e., no adjustment for landslide-road interactions) (**Input 6**). In this iteration, there were 60 direct road blockages, and (**Output 3**) 143 landslides within 50 m of the road and 270 landslides within 100 m of the road.

Figure 7.38 shows an example of a synthetic triggered landslide event inventory (**Output 1**) and the resulting direct road blockages caused by landslides (**Output 2**) for one out of one hundred iterations of LRIM for Su-Hua (northeast Taiwan) where the landslide density (L_D) = 0.35 LS km^{-2} (**Input 5**) and Road Landslide Protection set to 'standard' (i.e., no adjustment for landslide-road interactions) (**Input 6**). In this iteration, there were 3 direct road blockages, and (**Output 3**) 8 landslides within 50 m of the road and 19 landslides within 100 m of the road.

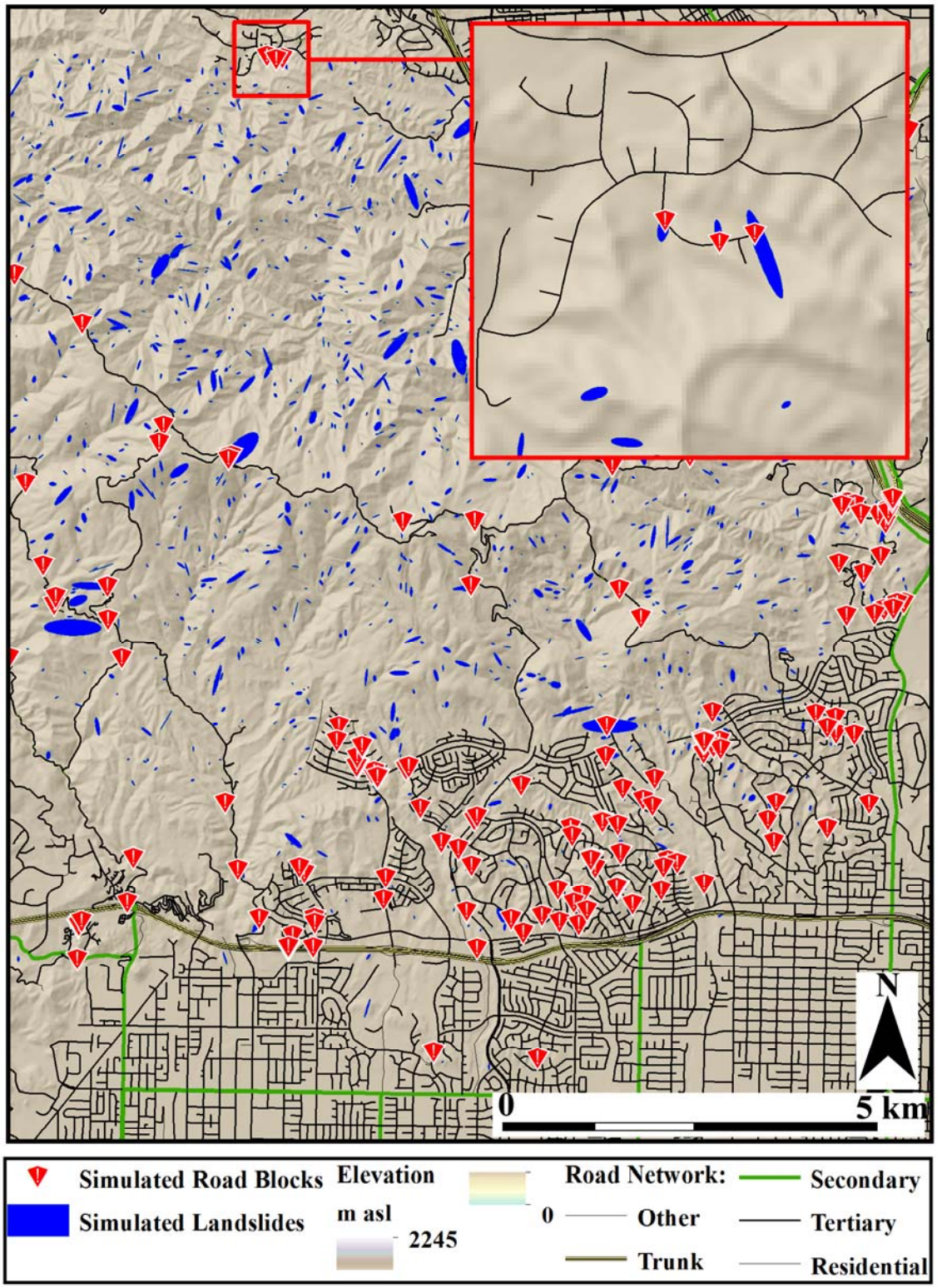


Figure 7.37 *Example Landslide-Road Impact Model (LRIM) output for the Oat Mountain (Northridge, USA) region showing a simulated landslide inventory (blue) and resulting road blockages across major and minor roads. This map shows results from one iteration of 100. In this simulation there are N_{LS} = 1,356 synthetic landslides resulting in 60 road blockages. The scale shown in the main figure is 5 km and the zoomed square is approximately 1 km×1 km.*

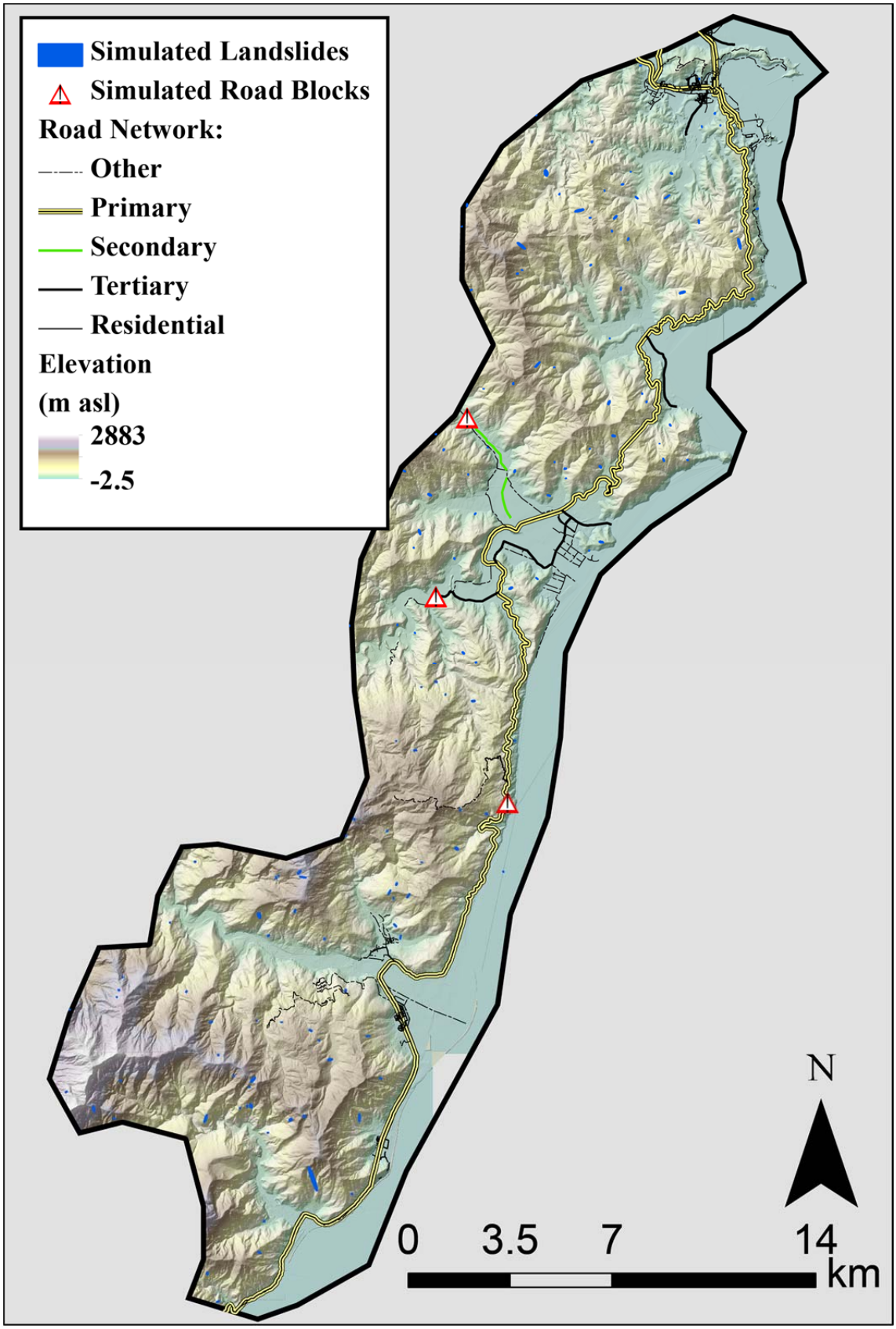


Figure 7.38 *Example Landslide-Road Impact Model (LRIM) output for the Su-Hua (Northeast Taiwan) region showing a simulated landslide inventory (blue) and resulting road blockages across major and minor roads. This map shows results from one iteration of 100. In this simulation there are* $N_{LS} = 149$ *synthetic landslides resulting in 3 road blockages.*

The number of road blockages (**Output 4**) from 100 iterations of the LRIM (configurations as given above) in both study sites are now examined. In Collazzone (Umbria, Central Italy), the number of direct road blockages (N_{BL}) ranges $24 \leq N_{BL} \leq 50$, with median 37 (equivalent to one road blockage every 4.1 km on average across the region). In Oat Mountain (Northridge, USA), the number of road blockages ranges $210 \leq N_{BL} \leq 327$, median 234 (equivalent to one road blockage every 3.3 km on average across the region). In Su-Hua (Northeast Taiwan), the number of road blockages ranges $3 \leq N_{BL} \leq 18$, median 8 (equivalent to one road blockage every 33.1 km on average across the region).

LRIM can be run for different landslide density scenarios to explore the relationship between the number of landslides triggered and the number of resulting road blockages. **Figure 7.39** shows, for Collazzone (Umbria, Central Italy) overall statistics (**Output 4**) of the number of road blockages for 8 different scenarios of landslide density (L_D), ranging from $0.5 \leq L_D \leq 7.5$ LS km^{-2}. For each density scenario, LRIM was run for 100 iterations. As landslide density increases, the number and the range of number of road blockages increases. Figures like this could be prepared in advance of a real landslide triggering event, so that if the number of landslides triggered can be quickly estimated following a triggering event, the number of expected road blockages (plus an estimate of uncertainty) can quickly be looked up to provide a rapid approximation of road network impact.

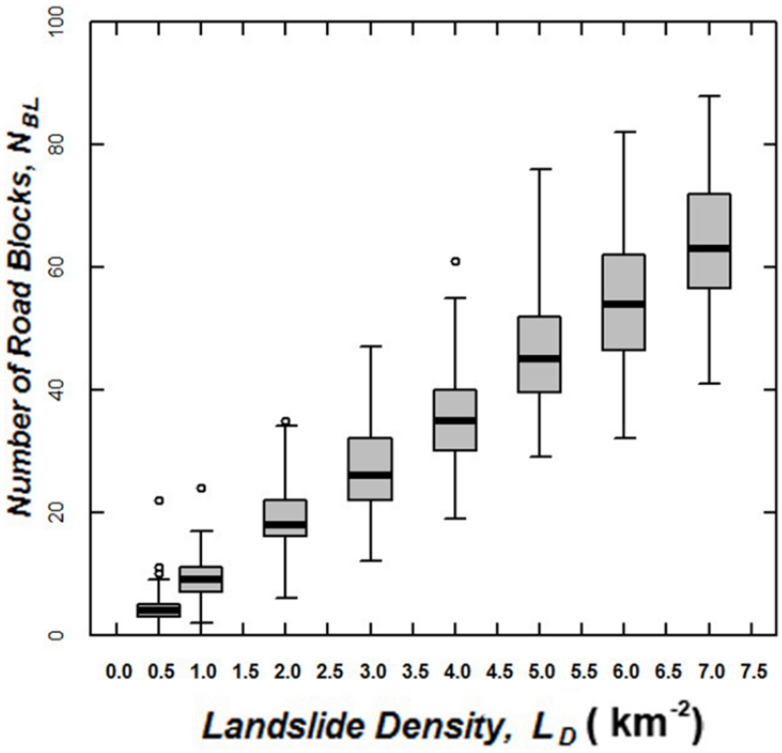


Figure 7.39 *Box-plot of relationship between landslide density and # road blockages for Collazzone (Italy).*

Further discussion of results, for the number of landslides in the corridors 50 m and 100 m from roads is discussed next in **Section 7.5.2**, along with model confrontation.

7.5.2 Model confrontation

For all three study sites, past triggered landslide event inventories are available to confront LRIM model outputs (**Sections 7.4.1.7** and **7.4.2.7**) and are shown in **Figure 7.40** (Collazzone, Umbria, Central Italy), **Figure 7.41** (Oat Mountain, Northridge, USA) and **Figure 7.42** (Su-Hua, Northeast Taiwan) along with resulting road blockages for these 'real' triggered events, and one iteration of a synthetic landslide inventory map created using LRIM.

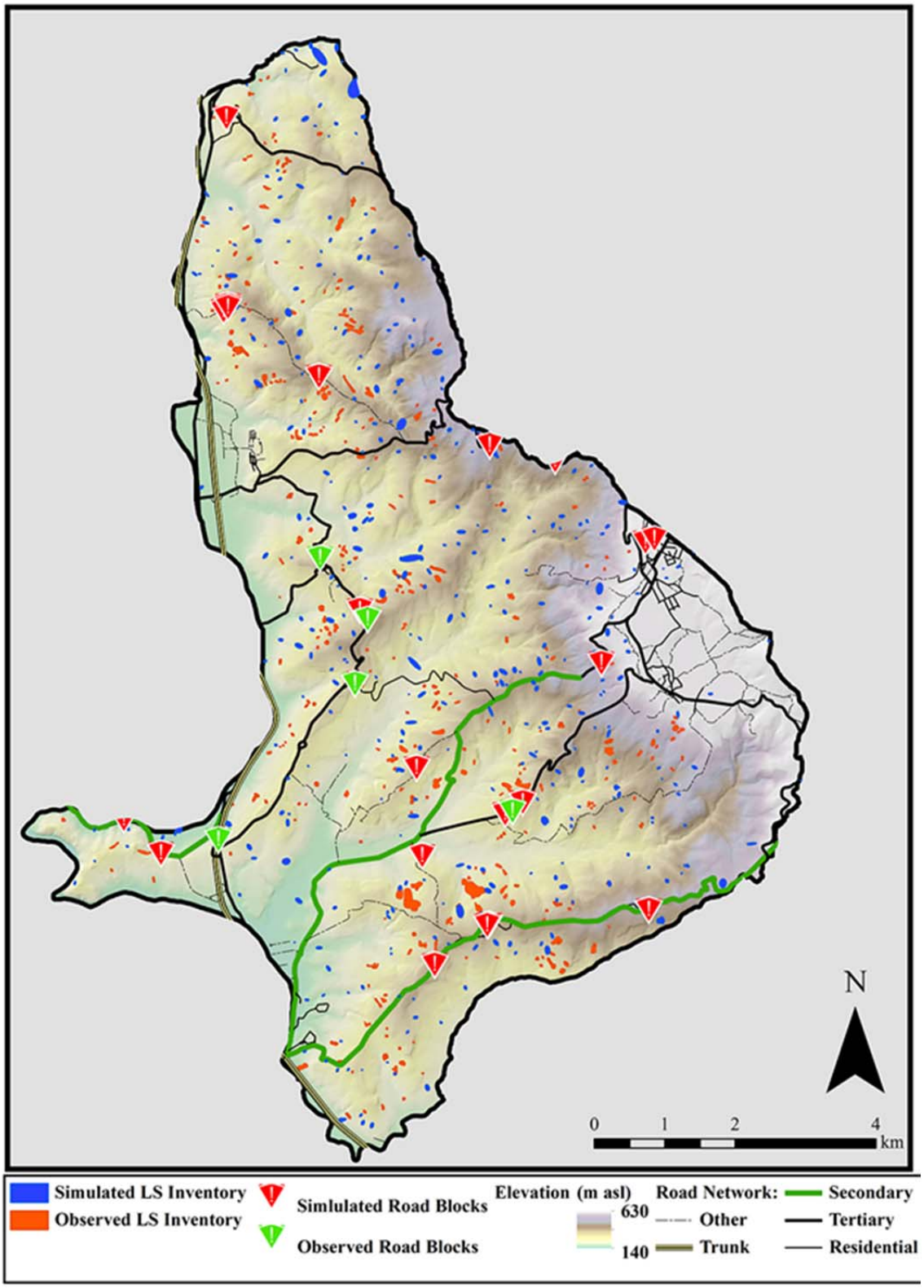


Figure 7.40 *Collazzone Region (Umbria, Central Italy). Observed triggered landslide event inventory from 1 January 1997 Umbria Region snowmelt triggered landslide event (Cardinali et al., 2000) compared to synthetic triggered landslide event inventory generated from one iteration of the Landslide-Road Impact Model (LRIM). Both the observed and simulated road blockage locations are shown (where a landslide polygon intersects a line). The scale in the figure is 4 km.*

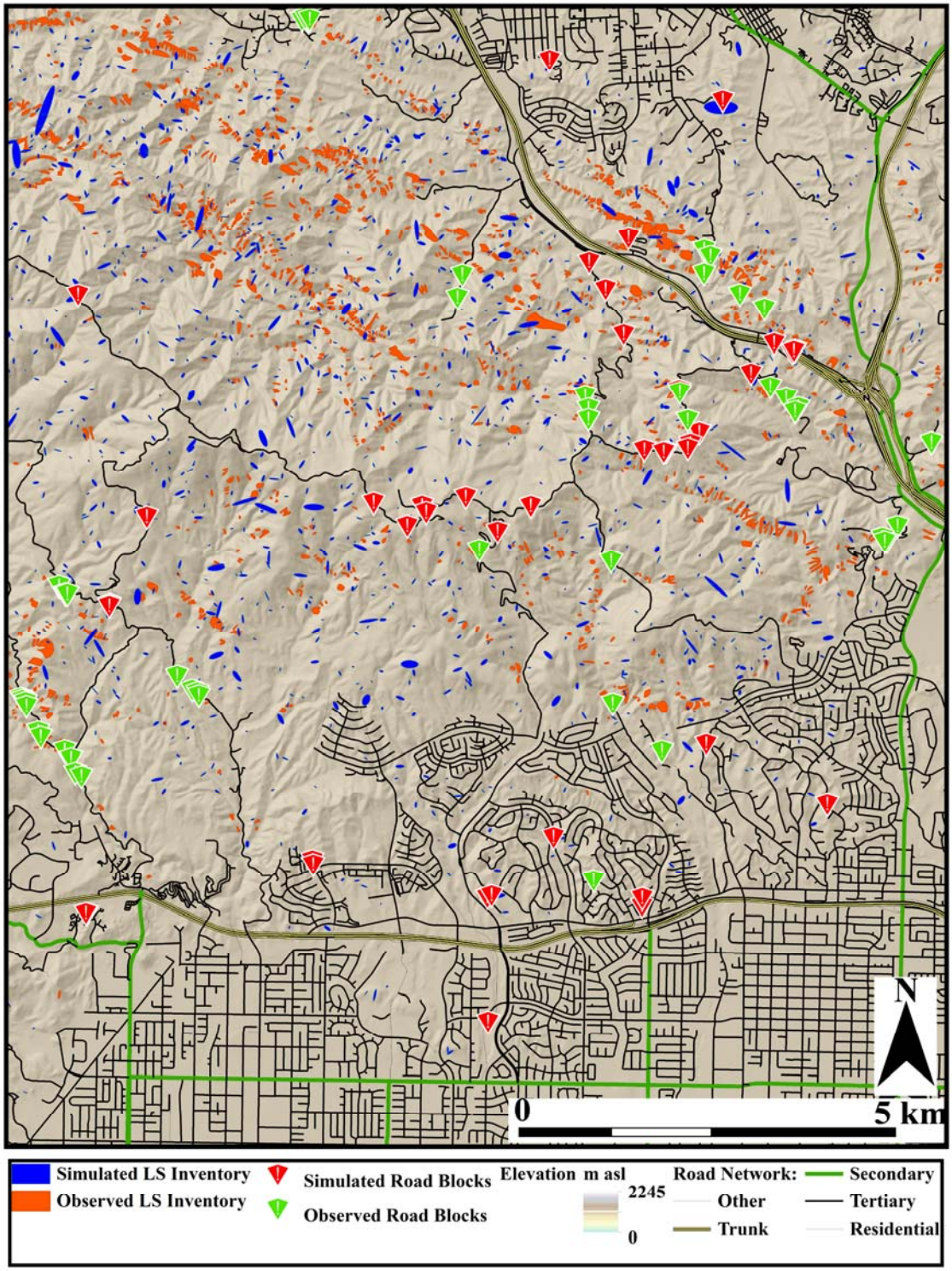


Figure 7.41 *Oat Mountain 7.5' Quadrangle, Northridge, USA (13.90 km × 11.49 km). Observed triggered landslide event inventory from 1994 Northridge Earthquake triggered landslide event (Harp and Jibson, 1995) compared to synthetic triggered landslide event inventory generated from one iteration of the Landslide-Road Impact Model (LRIM). Both the observed and simulated road blockage locations are shown (where a landslide polygon intersects a line). The scale in this figure is 5 km.*

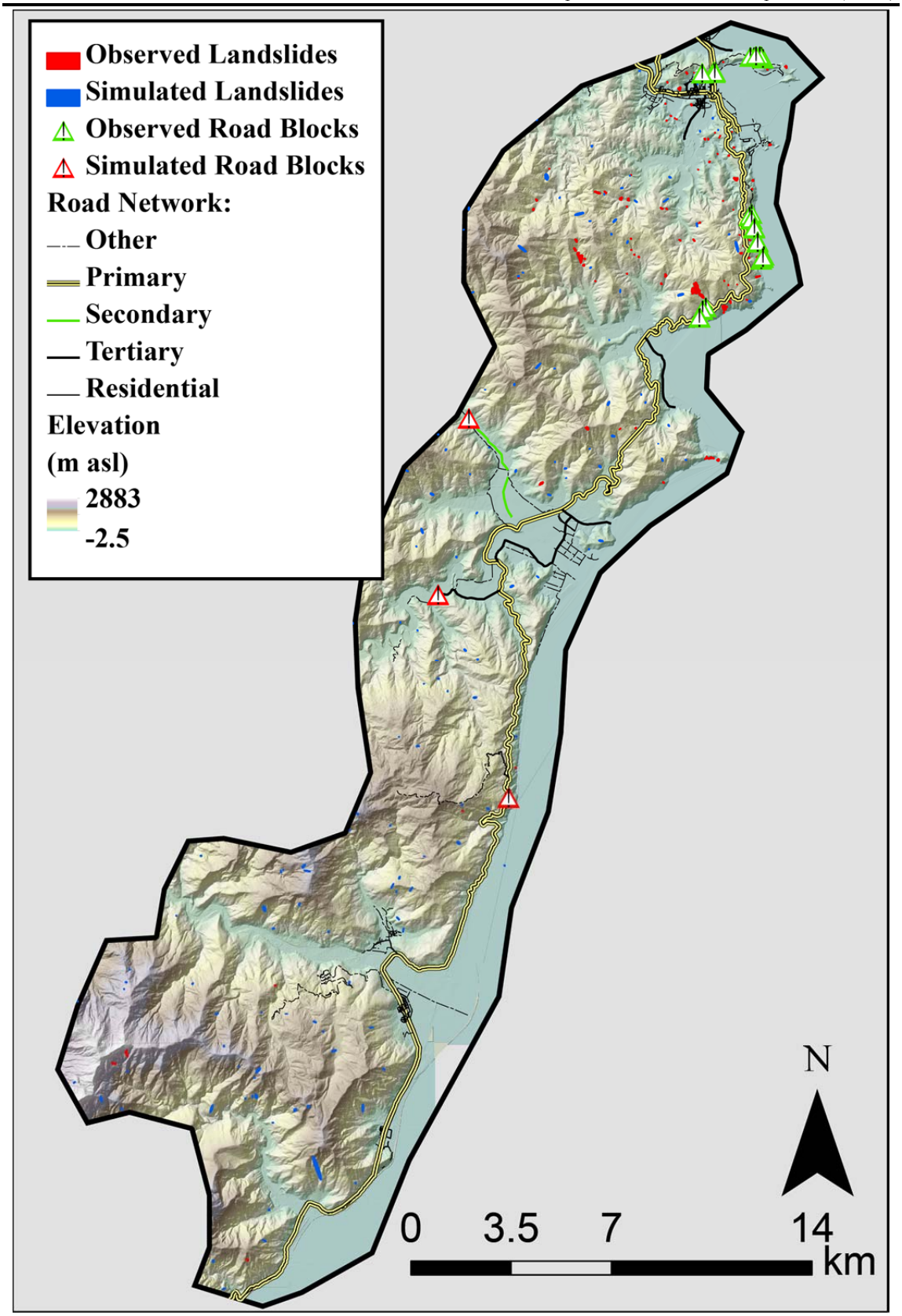


Figure 7.42 *Su-Hua road corridor study region, northeast Taiwan (13.90 km × 11.49 km). Observed triggered landslide event inventory from 2010 Typhoon Megi triggered landslide event (Chen, personal communication, 2014) compared to synthetic triggered landslide event inventory generated from one iteration of the Landslide-Road Impact Model (LRIM). Both the observed and simulated road blockage locations are shown (where a landslide polygon intersects a line).*

Figure 7.43 shows overall statistics (**Output 4**) of number of road blockages and landslides within 50 m and 100 m of the road compared to observations. For both Collazzone (Umbria, Central Italy) (**Figure 7.43A**) and Oat Mountain (Northridge, USA) (**Figure 7.43B**), LRIM over-predicts the number of direct road blockages by approximately a factor of 7 and 5 respectively. However, the model has greater skill in predicting the number of landslides within 50 m and 100 m of the road for these locations. For Su-Hua (Northeast Taiwan) (**Figure7.43C**), LRIM under-predicts the number of direct road blockages by approximately a factor of 2, and does not appear to significantly improve in forecasting the number of landslides within 50 m and 100 m of the road. **Figure 7.44** shows the modelled number of road blockages divided by the observed number ($N_{BL(Obs)}/N_{BL(Mod)}$): for both Collazzone and Northridge, as we look at an increasing buffer zone either size of the road network, the model skill improves. For Collazzone, the factor of overestimation decreases from × 7 for direct road blockages to × 2 for landslides within 100 m of the road, and a similar trend is observed for Oat Mountain. This alludes to the idea of landslide-road interactions at a finer spatial scale than the susceptibility map, discussed in **Section 7.3.2.6**, **Input 6**.

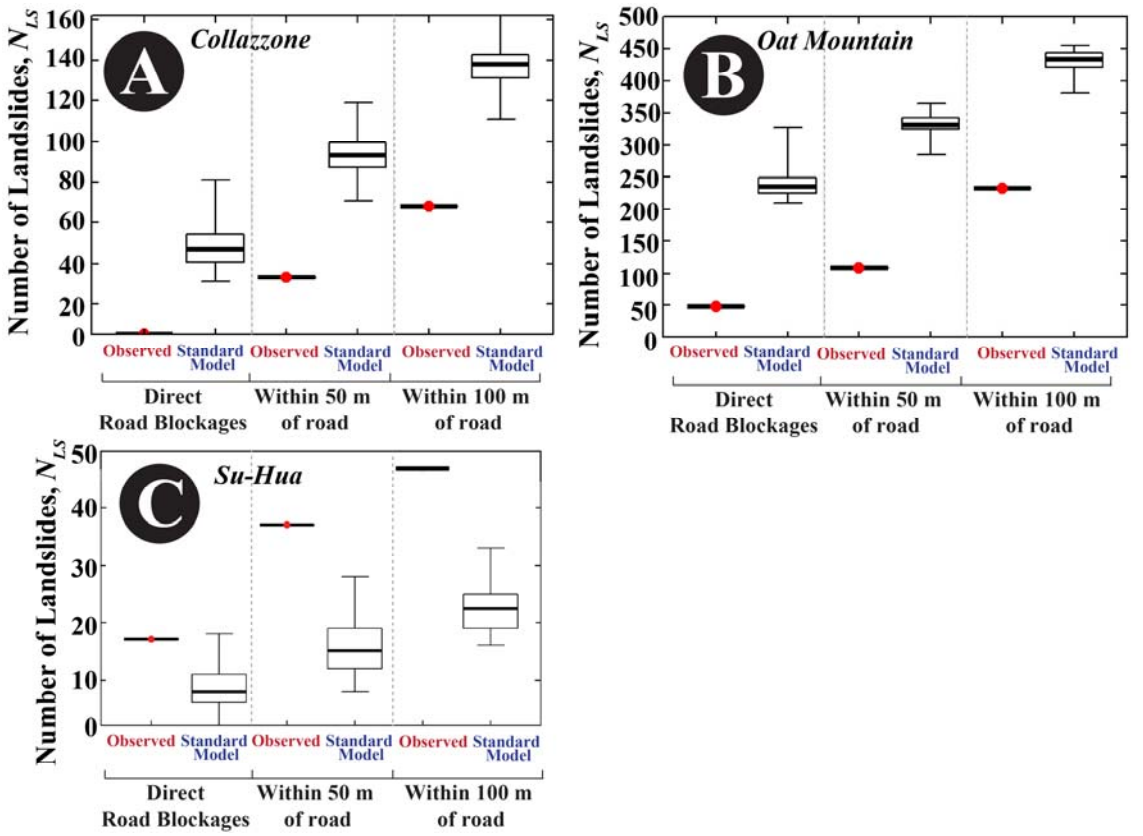


Figure 7.43 *Box-plots of observed (Obs) versus modelled (Mod) number of landslide road blockages (Direct) and number of landslides within 50 m and within 100 m of the road for* ***(A)*** *Collazzone (Umbria, Central Italy) and* ***(B)*** *Oat Mountain (Northridge, USA) using the standard Landslide-Road Impact Model (LRIM) configuration (not adjusting landslide susceptibility to take into account interactions between landslides and roads). Observed results come from triggered event inventories described in* ***Section 7.3****. LRIM landslide density parameter (****Input 5****) was set to match the average density of landslides from the observed events.*

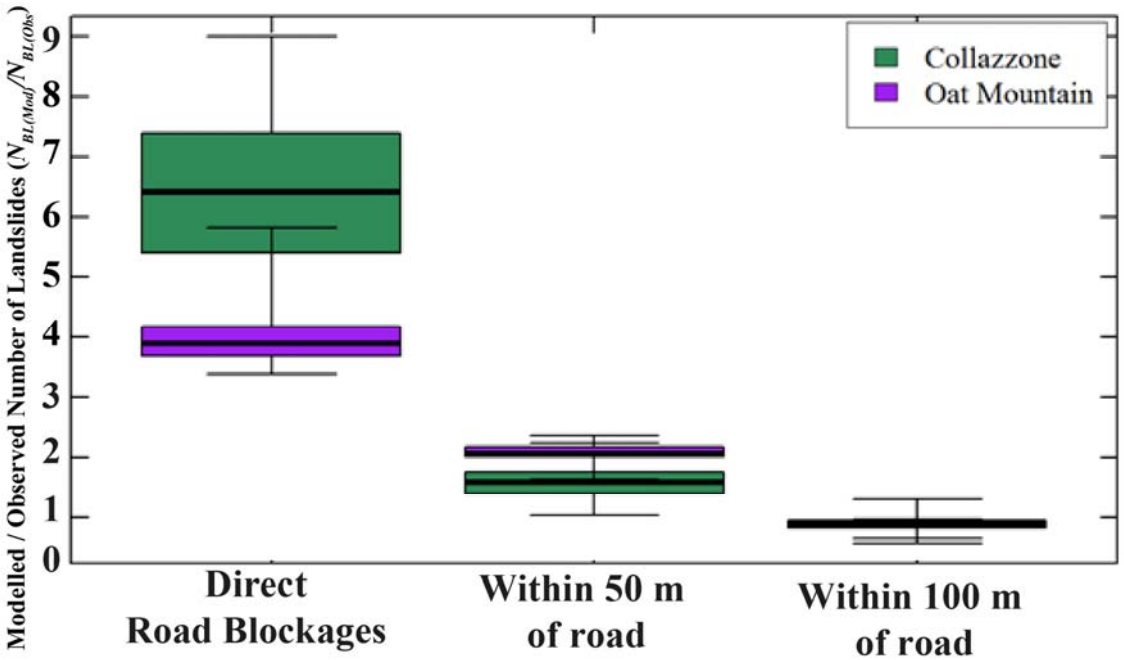


Figure 7.44 *Box-plot of the ratio (Modelled # road blockages)/(Observed # road blockages) for Collazzone (Umbria, Central Italy) and Oat Mountain (Northridge, USA) using the standard Landslide-Road Impact Model (LRIM) configuration (not adjusting landslide susceptibility to take into account interactions between landslides and roads). Observed results come from triggered event inventories described in* ***Section 7.4.1.7*** *and* ***7.4.2.7****. LRIM landslide density parameter (****Input 5****) was set to match the average density of landslides from the observed events.*

7.5.3 Experimenting with Road Landslide Protection Measures to improve LRIM output in line with observations

In this section, results are presented from experimenting with different parameters for LRIM **Input 6** (*Broad characteristics of Road Landslide Protection in a Region*). **Figure 7.45** shows overall statistics (**Output 4**) of number of the road blockages per iteration for four different configurations of LRIM for Collazzone (Umbria, Central Italy) (**Figure 7.45A**) and Oat Mountain (Northridge, USA) (**Figure 7.45B**) compared to observed values. The four model configurations (see also **Section 7.3.2.6)** and results are summarized as follows:

- **Standard Model:** LRIM run using standard **Inputs 1** to **4** (Road Map, Landslide Susceptibility Map, Study Area Boundary and Digital Elevation Model). **Input 5** Landslide density (L_D) set to density of observed inventory. **Input 6** *Broad Characteristics of Road Landslide Protection* is not implemented. For both Collazzone and Oat Mountain, LRIM over-predicts the number of road blockages and to a lesser extent, the number of landslides within 50 m and 100 m of the road.
- **Model Fixed *L/W*:** (only applied to Collazzone). LRIM run using standard **Inputs 1** to **4** (Road Map, Landslide Susceptibility Map, Study Area Boundary and Digital Elevation Model). **Input 5** Landslide density (L_D) set to density of observed inventory. **Input 6** *Broad Characteristics of Road Landslide Protection* is not

implemented. L/W of all ellipses was set to a fixed value of 1.8 (i.e., **Method 5** was not implemented). This was tested as no 'cut off' value is used for the inverse-gamma pdfs of L/W in LRIM, thus landslides may be unrealistically long and cause several road blockages. Although this did reduce the number of road blockages and landslides near roads slightly, it does not appear to be a significant factor.

- **Model Road Buffer 1:** LRIM run using standard **Inputs 1, 3** and **4** (Road Map, Map, Study Area Boundary and Digital Elevation Model). **Input 5** Landslide density (L_D) set to density of observed inventory. **Input 6** *Broad Characteristics of Road Landslide Protection* is set to Mode C (reduced landslide susceptibility within a buffer zone of the road). Buffer widths are configured using standard values of typical road widths for Europe and America, such that roads are buffered and landslide susceptibility reduced to 0 within these buffer zones. The width of the buffer corresponded to the width of that type of road (e.g., highways would be wider than residential roads). Examples of these general buffer widths for the USA are given in **Table 7.8**. For Collazzone (Umbria, Central Italy) this brings the direct number of road blockages and landslides within 50 m of the road closely in line with observations. For Oat Mountain (Northridge, USA) this does not significantly reduce the number of road blockages. Although such an approach would have to be applied carefully, and with knowledge of the local conditions, this could provide a rapid and generally applicable configuration for **Input 6** (minimal knowledge of local roads landslide protection measures required as typical road widths can be used).
- **Model Road Buffer 2:** LRIM run using standard **Inputs 1, 3** and **4** (Road Map, Map, Study Area Boundary and Digital Elevation Model). **Input 5** Landslide density (L_D) set to density of observed inventory. **Input 6** *Broad Characteristics of Road Landslide Protection* is set to Mode C (reduced landslide susceptibility within a buffer zone of the road). Road buffer widths are configured as, **Input 6** based on observed distances between landslides and roads in real inventories **(**described in **Section 7.3.2.6**). This requires some more detailed knowledge of local road landslide interactions (using a road map and observed landslide inventory) to establish distances between landslides and different types of roads. For both Collazzone and Oat Mountain (Northridge, USA) this brings the direct number of road blockages closely in line with observations. There is still some slight overestimation in number of landslides within 50 m and 100 m of the road.

Table 7.8 *Calculation of typical buffer widths for USA roads based on the number of lanes and standard lane widths given by (ref). We estimate the buffer width to be double that of the width of all lanes.*

Road Type	Number of Lanes	Total Road Width (m)(= Number of lanes × 3.66 m × 2)
Primary	8	58.56
Secondary	4	29.28
Tertiary	2	14.64
Minor	1	7.32

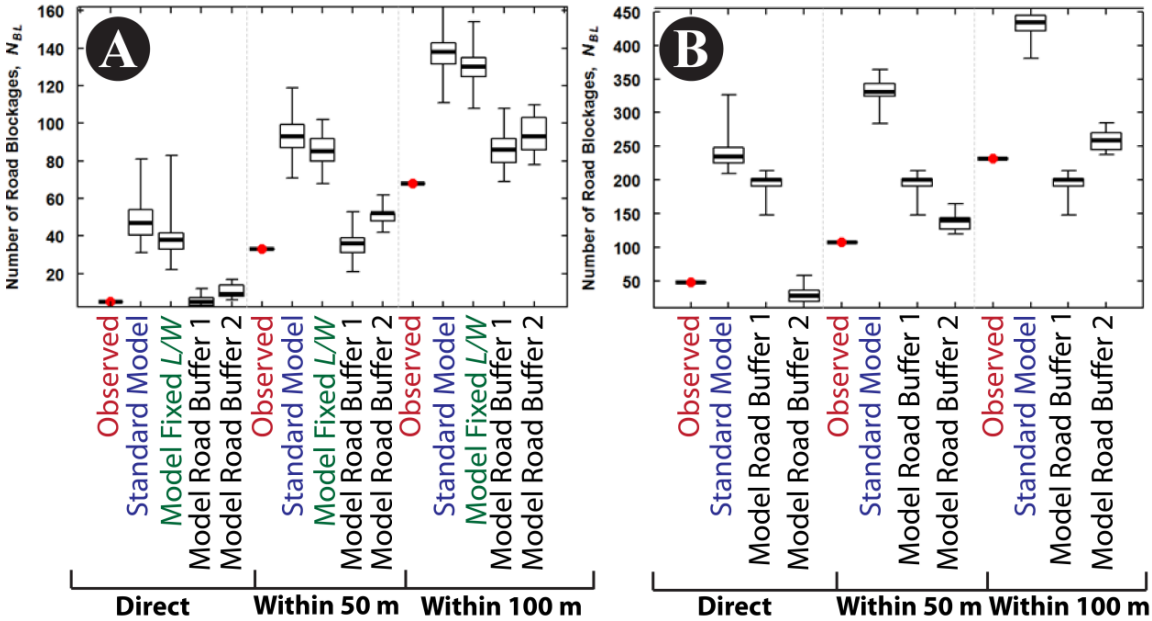


Figure 7.45 *Model versus observed number of road blockages for* ***(A)*** *Collazzone (Umbria, Central Italy, and* ***(B)*** *Oat Mountain (Northridge, USA). In this case the Landslide-Road Impact Model (LRIM) is configured to take into account interactions between landslides and roads (****Section 7.3.2.6****). On the x-axis are given number of landslides directly blocking the roads (****Direct****), and the number of landslides* ***within 50 m*** *and* ***within 100 m*** *of the road, for four model configurations (see also text): (i)* ***Standard Model*** *where there is no configured interaction set in the LRIM, (ii)* ***Model Fixed L/W*** *fixing all landslide elliptical shapes to L/W = 1.8. (iii)* ***Model Road Buffer 1*** *buffering the entirety of the road network by fixed widths depending on road type. (iv)* ***Model Road Buffer 2*** *randomly selecting portions of road network to buffer, the proportion of the total road length buffered and the width of the buffers depends on the road type. Within the buffer zones, the landslide susceptibility is decreased to 0. Note that the y-axes for the two inventories are different.*

For Taiwan, the standard LRIM configuration (**Input 6** set to **Mode A –** no adjustment of landslide susceptibility within road zones) slightly under-predicts the number of road blockages compared to observations of a triggered event in that region This suggests that for this site, landslide susceptibility may need to be increased in the corridor along roads (i.e., LRIM **Input 6** is set to **Mode C**) to more accurately reflect that road building in this location may increase the likelihood of landslides occurring through factors such as drainage and slope modification.

Overall, results from this section illustrate the importance of tuning the LRIM towards different 'systems' of road landslide protection. This has been done using detailed investigations of the road network and real triggered landslide event inventories specific to the study sites described in **Section 7.4**. However, as LRIM is applied to more locations and confronted with more real data, it may be possible to create more robust generalisations about different systems of road landslide protection.

Interaction between Landslide Susceptibility and Number of Road Blockages

To investigate further the link between landslide susceptibility and number of road blockages, the kappa-measure ($\varkappa$) methodology outlined in **Chapter 6** was used to measure the normalised number of road blockages per square kilometre grid cell ($\varkappa_{BL}$), relative to the (i) area of road buffered by 100 m normalised per grid cell ($\varkappa_{AR}$), (ii) median landslide susceptibility within that grid cell (S_{LS}), and (iii) number of landslide centroids normalised per square kilometre grid cell ($\varkappa_{LS}$),. In brief, the kappa-measure is calculated by:

1. *Creating a regular grid* of 1 km × 1 km grid cells that span the extent of the study area
2. *Overlaying the grid with the study area* to calculate what proportion of the cell is occupied by study area. If less than 25 % of the cell is covered by study area, the cell is rejected from further analysis.
3. *Overlaying the grid with a landslide susceptibility map* and calculating the median landslide susceptibility within each cell (S_{LS}).
4. *Overlaying the grid with a map of landslide centroids* and calculating the number of landslide centroids in each grid cell. If the grid cell is not fully covered by study area, the number of landslide centroids is normalised by study area to give the number of landslide centroids per square kilometre grid cell ($\varkappa_{LS}$).
5. *Overlaying the grid with a road network map* where the road line has been buffered by 100 m either side. The area of buffered road per square kilometre grid cell is calculated and normalised per square kilometre to give $\varkappa_{AR}$, as in Step 4.
6. *Overlaying the grid with a map of landslide-road blockages*. This is calculated by overlaying the landslide inventory with the buffered road network map described in Step 5. The number of 'road blockages' (landslides within 100 m either side of the road) is then normalised per square kilometre grid cell as in Step 4 to give $\varkappa_{BL}$.
7. *Moving the grid by 200 m* in the north direction, east direction and north and east directions in a moving window style approach and calculating Steps 2–6 for each cell in each movement of the grid.

Figure 7.46 shows a bubble plot of the kappa-measures calculated in Steps 4 – 6 for the Collazzone and Oat Mountain observed landslide inventories and LRIM model output for Collazzone and Oat Mountain. For each plot, the *x*-axis denotes the number of landslide centroids per cell, the *y*-axis denotes the number of 'road blockages' per cell (number of landslides within 100 m either side of the road), the size of the bubbles is proportional to the amount of road area within each cell and the colour of the bubbles represents median landslide susceptibility within that cell. For the Collazzone and Oat Mountain observed inventories (**Figures 7.46A** and **7.46C** respectively) we observe 'noisier' behaviour in the kappa-measures than we do in the LRIM model output (**Figures 7.46B** and **7.46D** respectively), where there is a clear positive relationship between landslide susceptibility (S_{LS}), area of road within a cell ($\varkappa_{AR}$), number of landslide centroids in that cell ($\varkappa_{LS}$), and resultant number of landslides within 100 m either side of the road ($\varkappa_{BL}$). This is similar to the results found in **Chapter 6** for the clustering of landslides, and suggests that the landslide susceptibility map does well model landslide-road interactions for Collazzone or Oat Mountain.

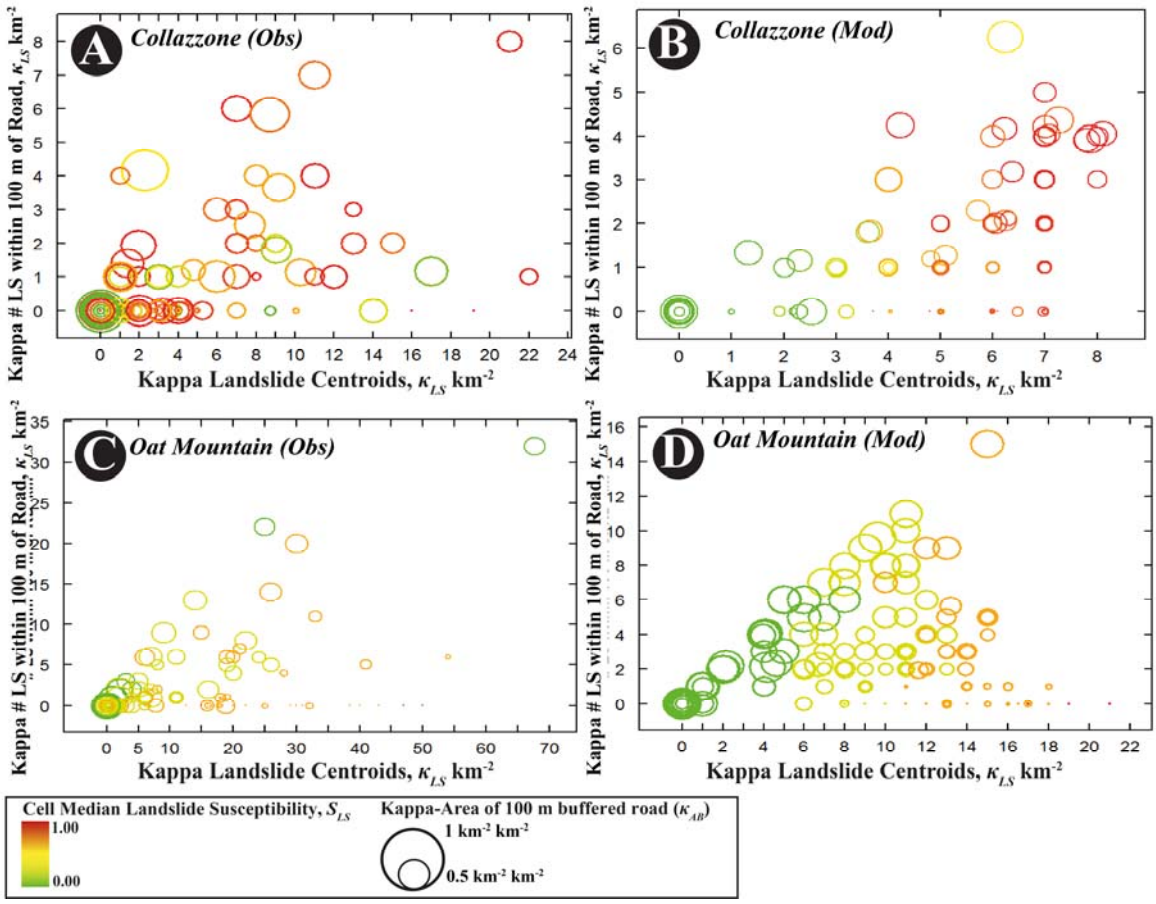


Figure 7.46 *Bubble plots showing the kappa-measures of number of observed landslide centroids per square kilometre ($\varkappa_{LS}$), kappa number of landslides within 100 m of the road per square kilometre ($\varkappa_{BL}$), kappa measure of the area of road buffered by 100 m per square kilometre ($\varkappa_{AR}$) and the median landslide susceptibility within that cell (S_{LS}) for* ***(A)*** *Collazzone, Central Italy and* ***(B)*** *Oat Mountain, California USA.*

In addition to in-built road protection measures reducing landslide occurrence in proximity to roads, we also discussed four other possible ways in which roads and landslides interact in the Collazzone region resulting in less road blockages than modelled by LRIM. The below four points are based on discussions in 2014 between myself, Bruce D. Malamud (King's College London), and colleagues from the CNR-IRPI (Perugia, Italy) who have been involved in LRIM development (Michele Santangelo, Ivan Marchesini, Fausto Guzzetti). These are:

- *The 'survival of the fittest' evolution of the road network*. Many of the roads in the Collazzone region are ancient (e.g., from Roman or pre-Roman times). We discussed the idea that over time, roads built through zones where landslides frequently occurred may have been abandoned (cf. Perkins, 1962). Over hundreds of years, this could lead to the evolution of the road network to travel through lowest susceptibility zones.
- *Road placements relative to typical landslide locations*. Meunier *et al.* (2008) have shown that earthquake triggered landslides tend to cluster near ridges due to the pattern of seismic shaking, whereas moisture triggered landslides tend to occur in the mid- to lower-portions of slopes (although this relationship does not appear to be 'generally applicable'). In Collazzone, roads tend to mainly be in valley bottoms or ridge crests, so could preferentially avoid the mid-slope regions where the majority of landslides occur, resulting in less road blockages. This concept was briefly investigated by calculating the percentage position up-slope for each observed landslide centroid and the centroids of 50 m segments of the road network for the Collazzone region. Results from this analysis shown in **Figure 7.47** illustrate that landslide centroids typically occur in the middle of slopes (**Figure 7.47A**) and roads tend to be in the upper or lower portions of slopes (**Figure 7.47B**).

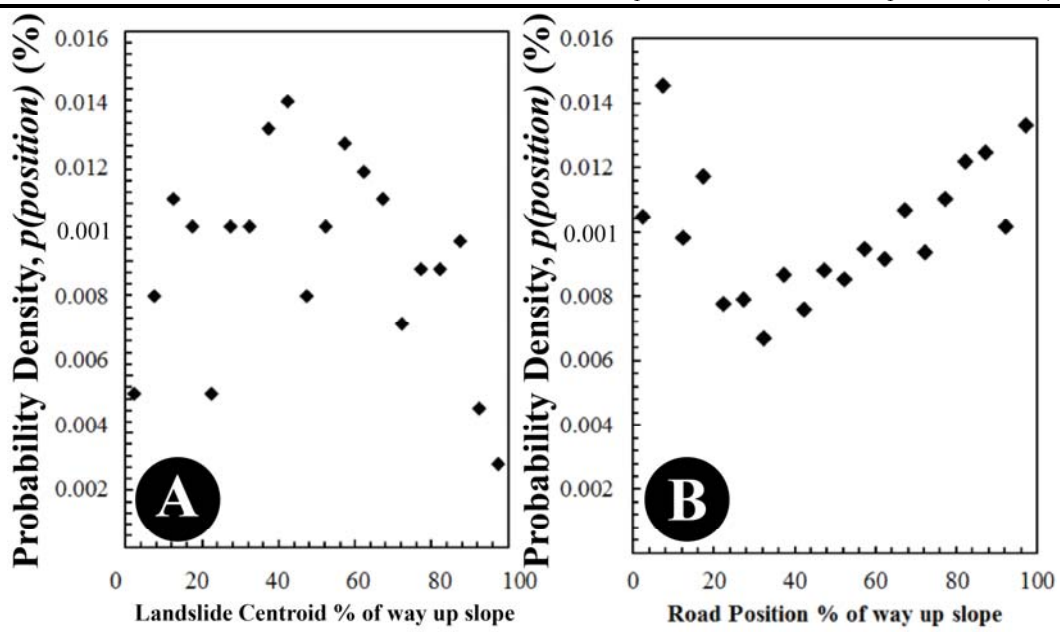


Figure 7.47 *Location up/down slope of* ***(A)*** *413 observed 1997 snowmelt triggered landslide centroids and* ***(B)*** *centroids of road network split into 50 m segments for the Collazzone region, Central Italy. % up slope is calculated by measuring the distances from the landslide/road centroid to the top and bottom of the slope unit divisions, as described in* ***Section 7.4.1.2.*** *Road map is 153 km of roads across the Collazzone region from open street map. Road was split into 50 m line segments and the centroid of each line segment calculated. % up slope of road centroid is calculated as per landslides.*

- *Roads acting as a barrier to landslides.* The road surface may actually stabilise a region, preventing shallow landslides from occurring, meaning that a larger, deeper mass would be required to overcome the stabilising effect of the road network. By splitting the Collazzone inventory into shallow and deep seated landslides and looking at proximity to roads (for all road types) (**Figure 7.48**), we see that this is probably not the case, as more shallow landslides occur in closer proximity to roads than deep seated landslides.

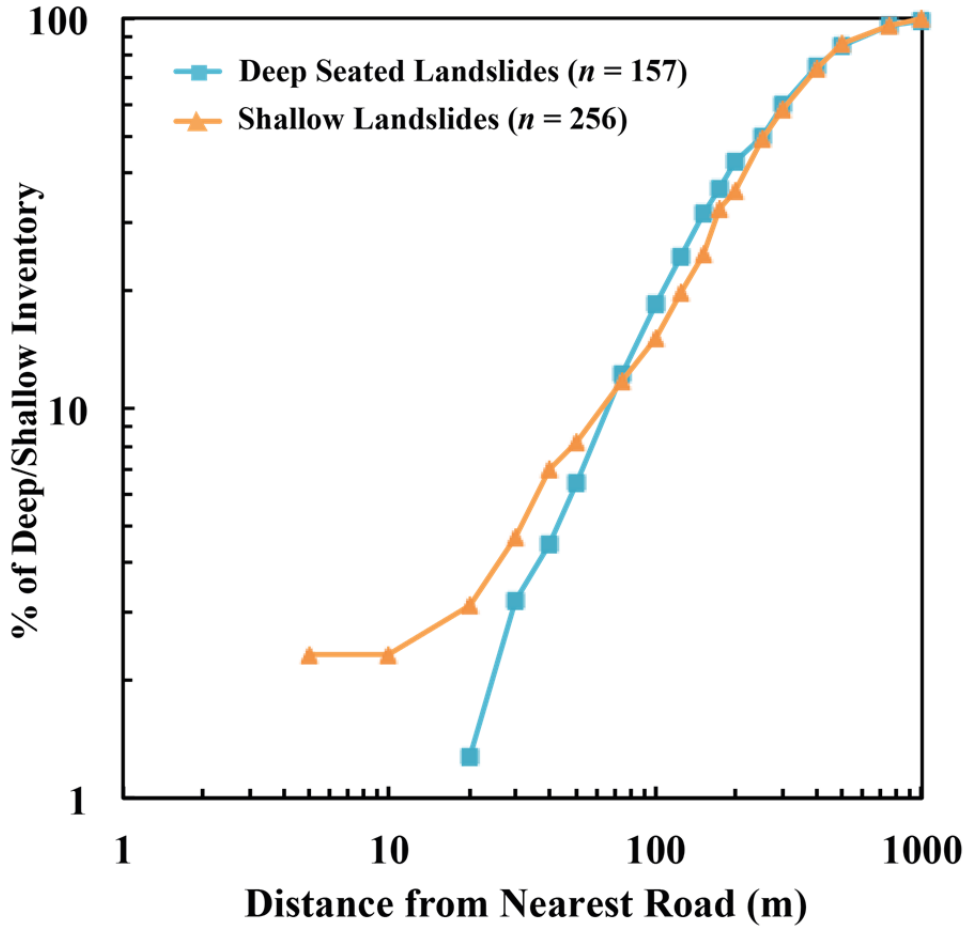


Figure 7.48 *Proportion of landslide inventory within a given distance of the nearest road for 157 deep seated and 256 shallow landslides triggered by 1997 snowmelt in Collazzone, Central Italy. Distances are binned using approximately logarithmically increasing bin sizes.*

- *Landslides cleared from roads before inventories are produced.* Although the Collazzone triggered landslide event inventory is considered 'high confidence' in terms of completeness, there were a number of days between the triggering event and the landslides being mapped (Cardinali *et al.* (2000) do not state the time lag between the triggering event and the date of the imagery being flown, but note that field surveys were carried out in the proceeding 10 days following the snowmelt). Between these dates, it is possible that some landslides would have been remediated (Baroni *et al.*, 2015), and that priority would be given to landslides causing human impact, such as those blocking roads. This could result in a preferential under-sampling of small, easy to remediate landslides with human impact. This highlights the importance of archive based inventory production and liaison with road managers to fully understand the incidence of landslide-road network impact.

Through the LRIM output confrontation against observed triggered landslide event inventories, we have shown that there appear to be some complexities in the way that landslides and roads interact, and this is not captured in the landslide susceptibility maps that we use. This can be approximately accounted for by adjusting the landslide susceptibility maps to increase or decrease the number of landslide centroids that are 'dropped' by LRIM within buffer zones of the road network. However, we have not been able to account for the physical mechanism behind this, and perhaps more data is required to understand these landslide road interactions.

7.5.4 Quantitative Road Network Analysis

Although not yet fully implemented within LRIM, it is possible to use graph theory based network analysis tools included in GRASS GIS to quantify the network disruption caused by a simulated triggered landslide event in terms of travel distance and connectivity between given points. A simple example is shown in **Figure 7.48**, where the impact on a journey between a pair of nodes ('Origin and 'Destination') is calculated. Normally, this is a 2.14 km journey (**Figure 7.48A**). When LRIM **Output 1** (a synthetic triggered landslide event inventory) is overlaid on the network (**Figure 7.48B**), the path between origin and destination is broken (**Figure 7.48C**). The new shortest path between the origin and destination nodes is then calculated using the GRASS GIS tool v.net.path, which is 8.94 km (**Figure 7.48D**).

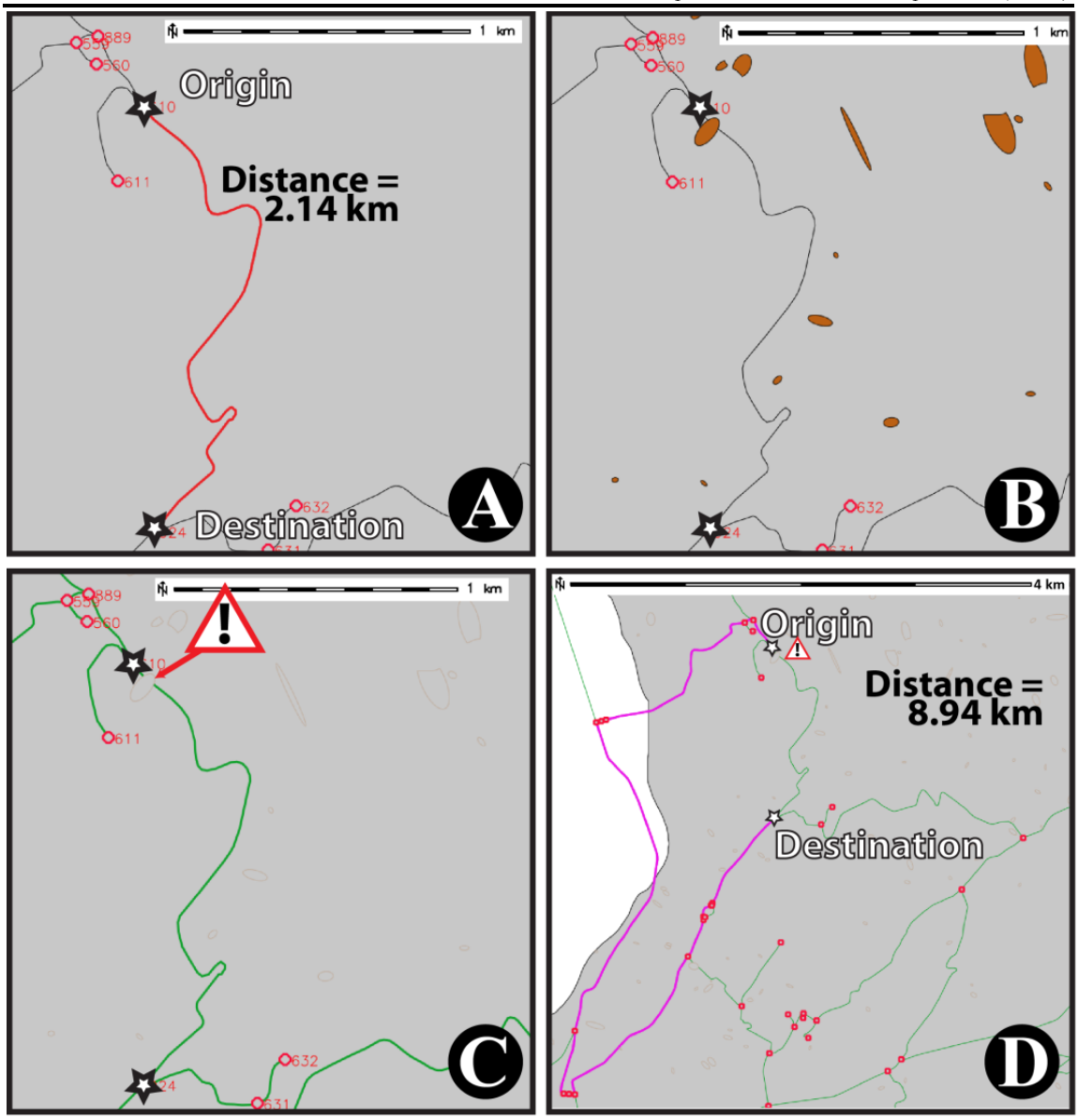


Figure 7.48 *Example of Network analysis incorporation into the GRASS GIS model. The scale in figures A, B and C is 1 km, and in D is 4 km.*

As discussed in **Section 7.2.3.2**, a number of network analysis indices exist to quantify the structure of a transportation network. Several of these can be applied to measure the change in road network structure following a landslide road network disruption from a triggered landslide event. For example, one of the simplest indices is the alpha measure (*a*) which is a ratio of the maximum possible number of circuits to the actual number of circuits in a graph:

$$\alpha = \frac{E-N+p}{\frac{N(N-1)}{2}\times(N-1)} \quad (7.2)$$

Where: α *= alpha measure*
E = number of edges
N = number of nodes
p = number of isolated subgraphs (if there are none, p = 1)
(Kansky, 1963).

If each node is connected by one edge to each other node in the graph, the graph is maximally connected and $\alpha = 1$. In real terms, this means that the road network has a high amount of redundancy; if one road becomes blocked, there will always be an alternative route between nodes (**Figure 7.49**, part i). For a more 'branched' road network where each node is only connected to a portion of all other nodes, α approaches 0, thus there is a higher probability that a node (or group of nodes) will become disconnected from the main graph if one road becomes blocked by a landslide (**Figure 7.49**, part ii and iii). **Table 7.9** describes some of the most widely-used indices and more recently developed indices that may be applicable to measuring the changes to a road network following a landslide triggering event.

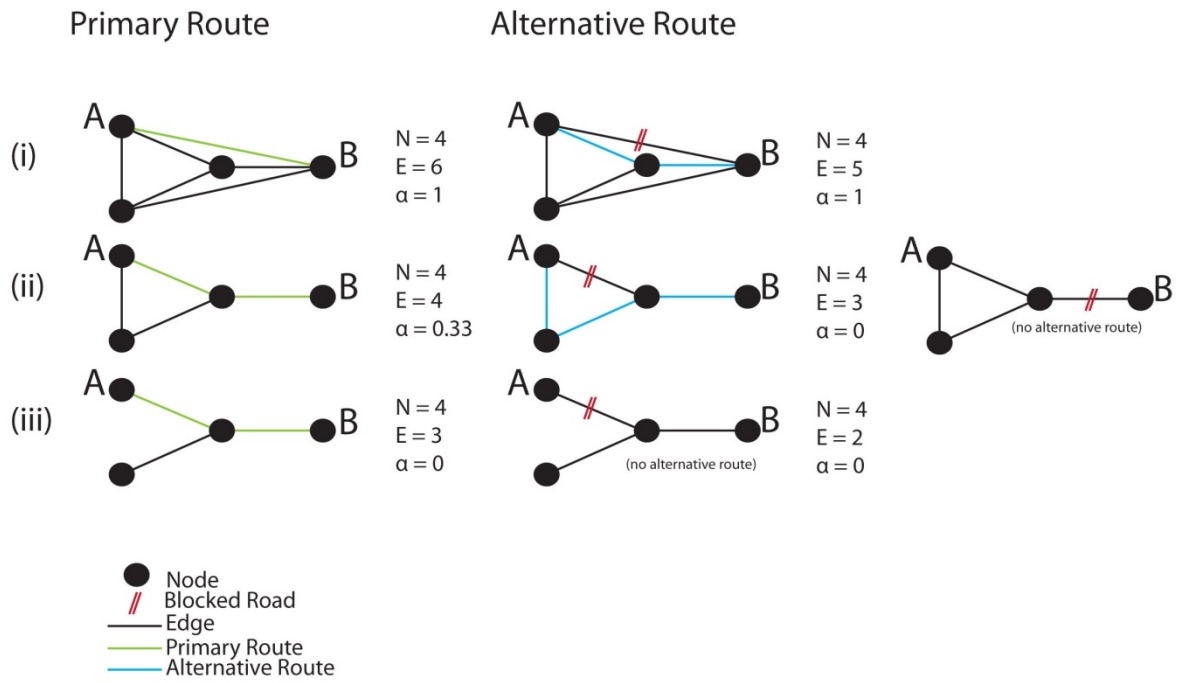


Figure 7.49 *Examples of varying α index and its effect when a road becomes blocked. In* ***Example i,*** *the graph is maximally connected so when any road becomes blocked, there is an alternative route between A and B (although the distance is longer). In* ***Example ii,*** *there is some redundancy; depending on which road is blocked there may be an alternative route between A and B. In* ***Example iii****, α = 0 thus if any road becomes blocked, one node will become isolated from the main graph. Developed from Kansky (1969).*

Table 7.9 *Key road network indices using graph theory in the literature*

Descriptor /Model	Formulae	Explanation	Reference
Beta Index (*β*)	$\beta_r = \frac{M_r}{N_r}$ *Where M_r is number of undirected edges for a region (r)* *N_r is number of nodes for a region*	Ratio of nodes to edges- gives an indication of the amount of redundancy/"treeness" of a network. Tends towards 1 for tree-like network, 2 for grid network	Haggett and Chorley (1969)
Alpha Index (*α*)	$\alpha_r = \frac{M_r - N_r + 1}{2N_r - 5}$ *Where notation as above*	Gives an indication of the amount of redundancy/"treeness" of a network. Close to 0 for trees and approaches 1 for triangular grids	Haggett and Chorley (1969)
Road Density (R_r)	$R_r = \frac{\sum \ell_r}{A_r}$ *Where ℓ is length of link within region r* *A is area of region r*	Indicates likelihood of alternative routes. If there are a relatively high number of roads in a region, it is more likely that there will be redundancy	Jenelius (2009)
Average Regional Traffic Flow ($\bar{f}_r$)	$\bar{f}_r = \frac{\sum_{k \in r} \ell_k f_k}{\sum_{k \in r} \ell_k}$ *Where notation is as above and f is traffic flow on link k*	The average traffic flow per link within a region. Indicates relative importance of region to economic activities.	Jenelius (2009)
Degree (*deg(v)*)	$\deg(v) = \sum E_v$ *Where E – edges entering vertex v*	Describes how well connected a node (vertex) is	Bollobàs (2001)

Descriptor /Model	Formulae	Explanation	Reference
Degree Distribution ($P_{(k)}$)	$P_{(k)} = \frac{n_k}{n}$ *Where n_k is number of nodes of degree k*	Examining the range and distribution of degrees of nodes gives insight into the network heterogeneity and homogeneity. Networks can be classified depending on what degree distribution they follow (e.g. small world, random, scale-free etc.)	Xie and Levinson (2007)
Entropy ($H(X)$)	$H(X) = \sum_{i=1}^{m} p_i \log_2(p_i)$ *Where m is number of subsets (e.g. highways, footpaths etc.) in system X* P_I *is proportion of agents (out of total system) in subset*	A second measure of network heterogeneity based upon functional groups. A homogeneous network (for instance, comprised entirely of motorways) has an entropy of 0.	Shannon (1948) in Xie and Levinson (2007)
Cyclomatic Number (μ)	$\mu = e - v + p$ *Where e is number of edges, v is number of vertices, p is number of sub-graphs*	Indicates the number of cycles (loops) within a network. A network with a high number of loops will have a high cyclomatic number. A fragmented network with few loops will have a low cyclomatic number.	Kansky (1963)
Efficiency ($E(G)$)	$E(G) = \frac{1}{N(N-1)} \sum_{i \neq j \in G} \varepsilon_{ij}$ *Where E(G) is network efficiency at time step G.* ε_{ij} *is efficiency of shortest path between node I and j (inverse of shortest distance)* *N is number of nodes*	Inverse of the shortest path length connecting all nodes to one another	Ash and Newth (2007)
Average shortest path length (L)	$L = \frac{1}{N(N-1)} \sum_{i=1}^{N} \sum_{j=i+1}^{N} L_{min}(i,j)$ *Where L_{min} is shortest path between I and j* *N is number of nodes*	Average of the shortest path connecting all nodes to one another	Ash and Newth (2007)
Gamma Index (γ)	$\gamma = \frac{e}{3\,(v-2)}$ *Where e is number of edges and v is number of vertices*	Ratio of the number of observed edges in a network to the maximum possible number of edges. Ranges from 0 (unconnected nodes) to 1 (maximally connected nodes where every node is connected to every node with a direct edge).	Kansky (1963)
Average Clustering Coefficient ($\bar{C}$)	$\bar{C} = \frac{1}{N} \sum_{i=1}^{N} \frac{2E_i}{k_i(k_i-1)}$ $\bar{C} = Average\,(\frac{Max\ no.\ links\ between\ neighbours)}{No.\ actual\ links\ between\ neighours)}$ *Where k_i is neighbours of node N_i* *E_i is maximum number of possible links between K_i neighbours*	Average amount of local clustering between links	Watts and Strogatz (1998)

Descriptor /Model	Formulae	Explanation	Reference
Network Diameter (δ)	$\delta = x \max y\, d(x,x)$ *Where d is topological length between nodes (number of edges)*	Maximum number of edges in the shortest path between two nodes in the network. A large network with a small diameter suggests a well-connected network.	Kansky (1963)
Pi index (π)	$\pi = \frac{C}{d}$ *Where C is total length of all edges within the network and D is length of network diameter (geographic units)*	A more connected graph will have a higher diameter as the maximum path length is a relatively small proportion of the total network length	Kansky (1963)
Inverse characteristic path length (L')	$L' = \frac{1}{n(n-1)} \sum_i \sum_{j \neq i} \frac{1}{d\,(v_i, v_j)}$ *Where $d(v_i,v_j)$ is the shortest path length connecting each node to each node If there is no connection between nodes, set to 0*	The average length of the shortest path between all nodes in a network	Duenas-Osorio *et al.* (2007)
Connectivity loss (C_L)	$C_L = 1 - \langle \frac{n_g^i}{n_g} \rangle_i$ *Where n_g are 'generating nodes' (e.g. electrical generators) and i is a substation fed by these generators*	Measures the average decrease in ability of nodes to receive flow from generating nodes. Most applicable to power and water networks but possibly could be adjusted for use with transportation networks.	Albert *et al.* (2004)
S-I Index	*($F(d_{ij})$) Where $d_{(ij)}$ = number of edges connecting any pair or nodes on the graph F = frequency*	Construction of a histogram based on the number of edges in the shortest path between each set of nodes. Analysis of skewness, modal values and moments of the distribution.	James *et al.* (1970)

This type of network analysis can be applied to all nodes in the network (a node is defined as where any two roads meet or a road ends) to quantify the impact of simultaneous road blockages caused by landslides on how easy it is to move about or access certain parts of a network. This type of analysis was applied to an output of 10 iterations of LRIM for Collazzone (Umbria, Central Italy) with landslide density = 5.13 *LS* km^{-2} and standard configuration for characteristics of road landslide protection. Results (**Table 7.10**) show that on average, over half of all possible routes between pairs of nodes are affected by the synthetic triggered landslide events. This type of analysis may be particularly useful for planning authorities in identifying locations particularly vulnerable to physical isolation and creating a more robust road network.

Table 7.10 *Example summary statistics of road network impact following a simulated triggered landslide event in the Collazzone region (Umbria, Central Italy) (landslide density = 5.13 LS km^{-2}, standard configuration for characteristics of road landslide protection). Averages of ten iterations of Landslide-Road Impact Model (LRIM).*

	Before Triggered Landslide Event	After Triggered Landslide Event
# of node pairs (all possible origins and destinations)	569,352	243,724
% of routes between all pairs of nodes closed	0 %	57 %
% of routes between all pairs of nodes disrupted	0 %	1 %
Average length of detour between pairs of nodes	N/A	1.4 km
Maximum length of detour between pairs of nodes	N/A	36.6 km

7.6 Summary and Discussion of Findings

7.6.1 Discussion

The literature reviewed in **Section 7.2.2** showed that most current methodologies to understand and forecast landslide-road network impact (i) do not consider multiple road blockages occurring as part of a triggered landslide event and (ii) often require large amounts of local data to parameterize models. Yet, as discussed in **Chapter 1**, many regions particularly vulnerable to large triggered landslide events where multiple roads may become blocked are low- to middle-income, and resources may not be available to implement effective tools to forecast landslide-road network impact. Some techniques that might potentially address this issue, include Monte-Carlo simulation (randomly selecting a given number of values from probability density functions and repeating this process hundreds of times) and graph-theory network analysis, both introduced in **Section 7.2.3**, where we examined literature and examples of natural hazard impact modelling from other disciplines.

Overall, this **Chapter 7** has shown the development of a relatively simple semi-stochastic Landslide-Road Impact Model (LRIM) to create 'synthetic' triggered landslide event inventory maps by exploiting 'generally applicable' statistical behaviour of triggered landslide event inventories (discussed in **Chapters 4** and **5**). The user is required to provide the following six inputs to the LRIM: road network, landslide susceptibility map, study area boundary, digital elevation, landslide density and broad characteristics of landslide-road protection measures. The LRIM works by randomly selecting landslide areas and shapes from statistical probability density functions. These landslides are semi-randomly 'dropped' over the study region, conditioned by the landslide susceptibility map. This creates a synthetic triggered landslide event inventory which can then be overlaid on the road

network. This process is repeated a hundred times in a Monte Carlo type simulation. The output is a probabilistic forecast of the regional road network impact from a particular triggered landslide event. For both Collazzone (Umbria, Central Italy) and Oat Mountain (Northridge, USA) test sites, the 'standard' LRIM configuration which has no change to susceptibility in corridors of roads compared to susceptibility in the region where the roads are placed, tends to *overestimate* the observed number of road blockages when comparing synthetic inventory simulations with an event that has occurred in each region. This is different than for Su-Hua (northeast Taiwan), where the standard LRIM configuration slightly *under estimates* the observed number of road blockages. When LRIM is configured to take into account some broad characteristics of road landslide protection in a region (by adjusting the landslide susceptibility map in corridors along the road network), the results for Collazzone (Umbria, Central Italy) and Oat Mountain (Northridge, USA) are closely in line with observations from real triggered landslide events. There are three modes possible to take into account landslide susceptibility in buffers near roads which can be set by the user depending on the prevalent susceptibility for their region.

In terms of creating a 'low data' input model, we have shown a model methodology that mainly uses generally applicable statistical behaviour of triggered landslide event inventories and globally available, open-access datasets. Three of the six LRIM user inputs, as follows, are relatively straightforward in terms of access to data and how they are used within LRIM:

Input 1: *A road network map.* We showed that the open-source, freely-available Open Street Map is sufficiently detailed in global coverage to use for most locations.

Input 3: *A study area boundary.* We have used three different types of study area boundaries for different locations (Collazzone, Italy = basin; Oat Mountain, USA = extent of available landslide susceptibility map; Su-Hua, Taiwan = manually drawn based on the extent of landslide inventory). The delineation of the study area extent does not appear to significantly affect the LRIM results (other than speed of processing time increasing for larger study areas).

Input 4: *A digital elevation model (DEM).* Because the DEM resolution is degraded to 100 m × 100 m resolution (**Method 7**), the globally available 90 m × 90 m shuttle radar topography mission elevation dataset (USGS, 2006) should be suitable for most locations.

The remaining three LRIM user inputs are somewhat more complicated and interact to affect the results of the LRIM modelled number of road blockages:

Input 2: *A landslide susceptibility map.* If a susceptibility map is not available, we showed how the freely available deterministic SHALSTAB model (**Section 7.3.2.2**) may be suitable to create a rough approximation of landslide susceptibility for a location. However, through analysis of the kappa-measure of normalised number of landslide centroids per km^2 ($\varkappa_{LS}$), normalised number of road blockages per km^2 ($\varkappa_{BI}$) and median cell landslide susceptibility (S_{LS}) in **Chapters 6** and **7**, we have illustrated the difficulties in using landslide susceptibility maps to forecast which parts of a landscape will fail during a particular triggered landslide event (i.e., some high susceptibility zones will fail multiple times and other high susceptibility zones will not fail at all. For all three locations, the susceptibility maps used were not able to recreate the observed spatial clustering of landslides nor the occurrence (or lack of occurrence for Collazzone and Oat Mountain) of landslides in proximity to the road network. We were able to improve the forecasting of number of road blockages by either increasing or decreasing landslide susceptibility in randomly selected buffer zones of the road network, which we believe reflects the effect of landslide road protection measures or road building techniques (e.g., lack of drainage) which may decrease or increase landslide susceptibility. Although making adjustments for increased/decreased landslide susceptibility near roads is not particularly deterministic or robust, it does reflect the reality that for both high- and low-income countries, where maps of landslide susceptibility in the immediate vicinity of roads are difficult to obtain.

Input 5: *Average landslide density (L_D) across the study region.* The user may choose any value of landslide density as an input to the LRIM, so that they might explore different scenarios of triggered landslide event magnitude (in terms of the numbers of landslides dropped across a region). However, as shown in **Chapter 6**, landslides cluster spatially, and so the actual density of landslides per km^2 will range considerably over a region in an observed triggered landslide event. The LRIM **Method 6** which semi-randomly drops landslides over the study region, conditioned by landslide susceptibility,

does not recreate the spatial scale of clustering observed in reality. **Chapter 6** showed some initial experiments that suggested the incorporation of clustering into the LRIM did not significantly improve model skill in forecasting number of road blockages, as in some iterations, landslides may cluster near to roads (causing several road blockages) and in other iterations, landslides may cluster away from roads (causing few road blockages), thus the average road network impact over all iterations will remain fairly similar. However, if the LRIM is applied to other setting types (e.g., to model the impact of landslides on buildings, agriculture, or different types of infrastructure), spatial clustering of landslides may have a greater effect on simulated impact results. One potential effect on spatial clustering of landslides that has not been investigated here is the trigger mechanism. Others (e.g., Densmore and Hovius, 2000; Crosta *et al.*, 2003; Meunier *et al.*, 2007; Meunier, 2008), have shown that landslide density varies across a region as a function of the type and magnitude of the triggering event (e.g., at a very basic level, a greater density of landslides will be observed at the earthquake epicentre than further distances). Potential developments of the LRIM could include the spatial location and extent of the triggering mechanism to determine where more or fewer landslides occur. However, this is dependent on the ability to record the location, extent and directional energy of the trigger. This may be relatively simple for earthquakes but more complex for hydro-meteorological landslide triggers that move across a landscape at a varied pace.

Input 6: *Broad characteristics of landslide-road protection measures*. We have discussed the issue of understanding how landslide-road protection measures may affect landslide susceptibility with regard to LRIM **Input 2** (landslide susceptibility map). Through confrontation of the LRIM outputs against an observed triggered landslide event inventory for Su-Hua (northeast Taiwan) the LRIM design appears to support a body of literature that suggests that landslide incidence may be increased in zones along road corridors due to issues of drainage and slope modification (e.g., Sidle and Ochiai, 2006; Miller and Burnett, 2007; Das *et al.*, 2012; Muenchow *et al.*, 2012). Here, a potentially novel contribution to the literature is a model whereby roads may either (i) not have an effect on landslide incidence; (ii) increase

landslide incidence due to poor drainage or road building; or (iii) decrease landslide incidence in buffer zones around the road network due to in-built road protection measures and road surfaces potentially inhibiting shallow landslide movement. Certainly, further research looking at landslide proximity and density near roads is needed to confirm this, but our method of randomly buffering road segments and reducing landslide susceptibility (LRIM **Input 5**) presents a way to approximate 'un-quantified' road protection measures across a region. Personal communication with both the British Geological Survey and Martin Klose (University of Vechta), who is investigating landslide impact on German road networks, suggests that because the age of many roads in high-income countries is > 100 years, it is very difficult to record the historical trajectory of road protection measures that have been implemented across a region. This illustrates how the low-data input approach here may also be useful in high-income countries when locally specific data is difficult to collect.

In conclusion, we have attempted to contribute to current knowledge of landslide-road interactions by showing a relatively simple method for simulating synthetic triggered landslide event inventories. Much like the United States HAZUS model (discussed in **Section 7.2.3.1**), the ability for the LRIM to forecast road network impact is more uncertain than very detailed, data-intensive studies. However, the methodology developed here is perhaps more realistic of the types of data and time available to forecast landslide-road network impact in many countries. Thus, the LRIM results could be considered a first approximation to guide further research within a region. The LRIM also (i) emphasises the importance of simultaneous road network blockages occurring as part of a triggered landslide event, which we believe has been under-examined in the current literature and (ii) demonstrates how generally applicable landslide statistical behaviour (discussed in **Chapters 4-6**) can be used in an applied sense. Until recently, much of the work on generally applicable landslide statistical behaviour seemed more concerned with the scientific accuracy and reasons behind results (e.g., Stark and Hovius, 1997; Malamud *et al.*, 2004a, Malamud *et al.,* 2004b, Katz and Ahranov, 2006; Stark and Guzzetti, 2009) rather than the real-world implications of these findings. The methodology developed here to create synthetic triggered landslide event inventory maps based on this statistical behaviour could now be used by others to simulate triggered landslide event impact on a variety of other systems (e.g., damage to buildings, risk of landslide-river damming, damage to

pipelines). The open-source model code will be made freely available in the near future, so that others may use and develop the tool.

7.6.2 Summary

- A methodology to build a Landslide-Road Impact Model (LRIM) has been developed. The LRIM creates synthetic triggered landslide event inventories from generally applicable statistical patterns of landslide area and shape and simulates the impacts of these triggered landslide event inventories on road networks across a region.
- The LRIM allows users to investigate triggered populations of landslides over a region and simultaneous road network impacts in a probabilistic manner and is thought to be generally applicable to regions susceptible to triggered low-mobility landslide events.
- The LRIM has been applied to three sites each with different geographic settings and typical landslide triggering mechanisms: Collazzone region (Umbria, Central Italy), Oat Mountain (Northridge, USA) and Su-Hua (Northeast Taiwan).
- The LRIM inputs are relatively minimal (compared to more typical landslide models that may require considerable local parameterisation, outlined in **Section 7.2**). User LRIM data inputs are: a road network map, a landslide susceptibility map, a study area boundary, a digital elevation model. The user is also required to set, for the region, average landslide density (L_D) and select a configuration for *broad characteristics of road landslide protection measures*.
- The LRIM works by randomly selecting landslide areas and ellipse length-to-width-ratios from statistical probability density functions. These landslides are semi-randomly 'dropped' over the study region, conditioned by the landslide susceptibility map. This creates a synthetic triggered landslide event inventory which can then be overlaid with the road network. This process is repeated a hundred times in a Monte Carlo type simulation.
- The LRIM was built in Python programming language using GRASS GIS, which is an open-source GIS software.
- Because of the open source approach and relatively minimal data requirements, the model is relatively low-cost to develop and operate. It is hoped this may encourage uptake of the LRIM methodology in low- to middle-income countries, where triggered landslide events can be particularly problematic.

The main findings for the Landslide-Road Impact Model (LRIM) are now summarised:

- We have calculated impact results using a hundred synthetic landslide inventory maps produced by the LRIM for three regions (Collazzone, Central Italy; Oat Mountain, California USA and Su-Hua, northeast Taiwan), and confronted this impact with that of a known triggered landslide event that has occurred in each region in the past.
- For both Collazzone (Umbria, Central Italy) and Oat Mountain (Northridge, USA) test sites, the standard configuration (i.e., no change to susceptibility in corridors of roads compared to susceptibility in the region where the roads are placed) of the LRIM tends to overestimate the number of road blockages by a factor of 2-7 when comparing synthetic inventory simulations with an event that has occurred in each region. For Su-Hua (northeast Taiwan), the standard

LRIM configuration tends to underestimate the observed number of road blockages by a factor of 2.

- One possible explanation for overestimation in the standard LRIM configuration is related to interactions between roads and landslides that are not captured at the scale of the landslide susceptibility map. These interactions may increase the likelihood of a landslide occurring on a road (e.g., steepening slopes) or reduce the likelihood (e.g., by incorporating road protection and slope stabilisation measures).
- When the LRIM is configured to take into account some broad characteristics of road landslide protection in a region (by randomly buffering segments of the road network and adjusting the landslide susceptibility map in these buffer zones), the results for Collazzone (Umbria, Central Italy) and Oat Mountain (Northridge, USA) are closely in line with observations from real triggered landslide events.
- It is possible to exploit graph theory network analysis tools available in GRASS GIS to provide quantitative estimates of road network disruption in terms of travel distance and network connectivity.
- This Landslide-Road Impact Model presents an open-source low-data methodology for the stochastic simulation of potential simultaneous road network impacts.

Chapter 8: Summary and Conclusions

8.1 Introduction

As introduced in **Chapter 1** of this thesis, the estimation of landslide hazard and risk according to Guzzetti *et al.* (1999) is often limited by:

1. *Poor or partial records* of past landslide events.
2. *Access to quality datasets* of local conditions.
3. *Selection of appropriate techniques and tools* to calculate landslide hazard.

These three limitations are united by the issue of data (quality, availability and appropriateness of methods of analysis for the data available). This thesis has developed methodologies and tools that have attempted to contribute to answering the question "*how can we work around the issue of low data confidence for understanding landslide risk and impact?*". The majority of this chapter (**Section 8.2**) provides a summary of how **Chapters 2** to **7** have each answered parts of the original four research questions outlined in **Chapter 1**, including a discussion of where I believe original contributions have been made in background and primary research chapters to the knowledge on this topic. The final section of this chapter (**Section 8.3**) concludes with brief remarks.

8.2 Summary of Chapters

This thesis aimed to contribute to our knowledge of "*how we can work around the issue of low data availability for understanding landslide risk and impact*". This was done through four primary research questions (**RQ.**) across three different but inter-linked thesis parts:

Thesis Part I: *Landslide Inventories and Completeness* (**Chapters 2** and **3**)

> **RQ. A.** Is it possible to obtain richer information about the occurrence and impacts of landslides in Great Britain from regional newspapers?

Thesis Part II: *Statistical Models of Landslide Size, Shape and Spatial Clustering* (**Chapters 4**, **5** and **6**)

> **RQ. B.** What shape can landslides be modelled as, how does this shape vary, and is this 'generally applicable' behaviour across different landslide inventories from different locations?

RQ. C. How can the spatial clustering of landslides in triggered event inventories be measured and is this 'generally applicable' behaviour across different landslide inventories from different locations?

Thesis Part III: *Development of a Landslide-Road Impact Model* (**Chapter 7**)

RQ. D. Is it possible to create a 'generally applicable' model of triggered landslide event impact upon regional road networks using the statistical behaviour of landslide size and shape?

Chapters 2 to **7** are now summarised, with emphasis on where we believe we have made original contributions to the knowledge.

8.2.1 Thesis Part I: Landslide Inventories and Completeness

Chapter 2 of the thesis was a background review of literature and concepts, establishing the importance of understanding what type of landslide inventories are used, how different types of inventories may have different kinds of uncertainties in terms of what is recorded and how these uncertainties can impact the results of analysis. Particular emphasis was placed on the concept of spatial persistence of landslides: that the geomorphic signature of smaller landslides tend to be erased from the landscape more rapidly than larger landslides, which can often create a bias towards recording proportionally more large landslides in an inventory. **Chapter 2** explained that by using triggered landslide event inventories that are systematically produced at a fine scale soon after a triggered landslide event occurs, the inventory is more likely to reflect the full range of landslide types, shapes and sizes encountered across the landscape without bias. Harp *et al.* (2011)'s criteria for substantial completeness of a triggered landslide event inventory was introduced, which established a baseline for the analysis of triggered landslide event inventories performed in **Part II** of this thesis. Although, as Malamud *et al.* (2004a) note, an inventory of landslides will never be fully complete.

Chapter 3 of this thesis was a primary research chapter aiming to answer **RQ A** "*Is it possible to obtain richer information about the occurrence and impacts of landslides in Great Britain from regional newspapers?*". **Chapter 3** developed a method to address the issue of landslide database completeness through construction and application of a robust Boolean search to a digital archive of 568 UK regional newspapers. This was done to identify more records and enhance the richness of information about landslides to supplement the *National Landslide Database* (NLD) of Great Britain. This chapter acknowledged from the outset that

application of this method will not produce a complete inventory of landslides, but is a way to generate more and richer records of landslides. The numbers and types of landslides identified from newspapers can also be compared to the current NLD records to explore current NLD database completeness.

Although searching newspapers for information about landslides is not a new technique (e.g., Guzzetti *et al.*,, 1994; Kirschbaum *et al.*, 2010; Petley, 2012) as far as we are aware, our study is one of the few examples in the literature where the search terms and strategy are clearly outlined, which illustrates the robustness and replicability of this methodology, and an original contribution to knowledge. We hope that this method will now be used by others to add richness to other landslide databases. Through the application of the Boolean search to two years (2006, 2012) of articles published in 568 regional newspapers, we identified 112 additional landslide events not currently recorded in the NLD, representing a 120 % and 40 % increase in number of landslide records for 2006 and 2012 respectively. We also showed that we were able to enhance landslide information in the NLD by obtaining further information about landslide impact for 60–90 % of landslide events identified from newspaper articles, and argued that the current structure for recording impact in the NLD is does not capture well the variety of impacts that landslides cause. This difficulty in recording landslide impact was also discussed in **Chapter 7**, where we found that several studies underestimated landslide-road impact due to partial or incomplete records of past events. Spatial and temporal patterns of additional landslides identified from newspaper articles were broadly in line with those currently in the NLD. This consistency between our findings and existing records in the NLD suggests that although the NLD is not a complete record of landsliding in Great Britain, it does not exhibit a strong spatial or temporal bias. The 40 % (2012) and 120 % (2006) increase in number of landslides recorded and the 60–90 % of the landslide records with enhanced impact information shows that it is possible to obtain richer information about landslides in Great Britain through the systematic search of regional newspaper archives (**RQ. A**).

8.2.2 Part II: Statistical Models of Landslide Size, Shape and Spatial Clustering

Part II of this thesis moved from discussing the completeness of landslide inventories and databases to the statistical analysis of substantially complete triggered landslide event inventories.

Chapter 4 was primarily a background literature and concepts review of statistical patterns of landslide areas in triggered landslide event inventories. We first briefly explained the concepts of frequency and probability density as ways of calculating the relative likelihood of observing an event of a given magnitude. We then introduced key papers by Stark and Hovius (2001) and Malamud *et al.* (2004a) which both showed that landslide probability density is (i) relatively low for small landslide areas, (ii) has a characteristic 'rollover' (location of maximum probability density) for medium landslide areas and (iii) has an inverse power law decay for medium to large landslide areas. Although the authors use different probability density functions to describe this behaviour (the double Pareto and inverse Gamma distributions, respectively), the overall trend is the same, and has important implications for understanding of landslide risk as part of triggered landslide events, with the landslide area distributions strongly non-normally distributed. Through the analysis of three large, substantially complete triggered landslide event inventories, Malamud *et al.* (2004a) showed that there is a generally applicable statistical behaviour for the frequency-area distribution of landslides in a triggered landslide event, regardless of location or landslide triggering mechanism. It is not fully understood why this behaviour is generally applicable, but it appears to be physically based (rather than an outcome of uncertain or biased data) and recent theories seem to be converging around the themes of landslide depth, soil cohesion and friction as the principal controls of landslide area probability density distribution. The finding of this 'generally applicable' statistical behaviour provided the motivation for research in **Chapters 5** and **6**, discussed next, where shape and spatial clustering in triggered landslide event inventories were explored as to 'generally applicable' behaviour.

Chapter 5 was a primary research chapter aiming to answer **RQ. B** "*What shape can landslides be modelled as, how does this shape vary, and is this 'generally applicable' behaviour across different landslide inventories from different locations?*". **Chapter 5** presented analyses of the shape of individual landslides using two large, substantially complete triggered landslide event inventories (as defined in **Chapter 2**) and the additional analyses of five 'lower confidence' triggered landslide event inventories which ranged in location, triggering mechanism and number of landslides recorded. We first presented a review of existing literature showing that an ellipse is a reasonable approximation of landslide shape, and that landslide length to width ratio (L/W) typically ranges $0.28 < L/W < 10.0$. We then presented four methods for approximating the shape of landslide polygons to ellipses, and a measure developed by Lombardo (2014) to test the goodness of fit of an ellipse to an irregular polygon. We found

that fitting a convex hull (CH) to each landslide polygon, calculating an ellipse with the same area (A_{CH}) and perimeter (P_{CH}) of the convex hull, and then scaling this ellipse to match the original landslide area (A_L) was a good method to approximate the shape of landslides to ellipses. The methodology we have developed to rapidly and systematically measure L/W of large numbers of landslides in a GIS presents a useful supplement to the literature where the majority of existing studies of measured landslide L and W in the field on small samples of landslides.

We found that generally > 80 % of low-mobility landslide polygons are reasonably well approximated by an ellipse; this is not true for high-mobility landslides (e.g., long sinuous debris flows). This partially answers **RQ. B** that low-mobility landslides are typically approximately elliptical. We believe that this systematic review of the shape of landslides across seven triggered landslide event inventories (totalling 23,276 landslides) is a useful confirmation of anecdotal evidence in the existing literature that landslides are approximately elliptical (e.g., Marchesini *et al.*, 2009; Martha *et al.*, 2010). Of landslides that are well-modelled by an ellipse, L/W ranges ~$1.2 < L/W \sim 15$. This is slightly broader than the range of L/W identified in the literature (max $L/W = 10$). This may be due to the inclusion of both source area and run-out in our analyses (as these are not separated in the inventories we use). However, we believe that the quantification of the shape of the full extent of a landslide (including run-out) is useful for practical applications, where civil protection agencies wish to forecast the full extent of land affected by an individual landslide (i.e., including run out).

Chapter 5 also presented a methodology to split landslides into logarithmically increasing landslide area (A_L) categories and analyse the distribution of L/W within each area category, showing that an inverse Gamma probability density function is a robust model of L/W distribution for each area category. The inverse Gamma pdf has also been used to model the distribution of landslide areas (as discussed in **Chapter 4**), which suggests potential similar controls of landslide area and shape, and potential interactions.

The inverse Gamma pdf models a low probability of observing small values of L/W. There is a rollover for medium values of L/W and then an inverse power-law decay for large values of L/W. The parameters describing the shape of the inverse Gamma pdf vary with landslide area category, which contributes to answering **RQ. B** that landslide shape varies with area. This finding appears to be in opposition to existing 'rules of thumb' in the

literature (e.g., that suggest that landslide shape remains approximately constant with area, Guthrie *et al.*, 2008).

For the six moisture-triggered landslide inventories analysed (Italy (× 3), Guatemala, El Salvador and Taiwan) broadly, the location of the rollover and the tail of the decay suggest that as we approach larger landslide areas, landslide L/W tends towards higher values (landslides are longer and thinner). Whereas, for the one earthquake triggered inventory analysed (Northridge, USA), the opposite trend is seen. However, the five 'lower confidence' landslide inventories introduced considerable 'noise' in the overall trends found in terms of wider variability in inverse Gamma pdf parameter values fit to each dataset.. We are unsure whether this 'noise' this is simply a result of smaller sample sizes and potentially biased data, or actually suggests that landslide shape behaviour differs between different locations. This partially answers **RQ. B** that the statistical behaviour of landslide shape may be semi- 'generally applicable' if separated by triggering mechanism. This also opens up potential lines of enquiry as to why the statistical behaviour of landslide area is more definitively 'generally applicable' than the statistical behaviour of landslide shape, and what processes may govern this. For the modelling of synthetic triggered landslide event inventories performed in **Chapter 7**, we assume that this behaviour of L/W is in fact, universal, although the analysis of additional 'high confidence' triggered landslide event inventories is required to confirm this.

Chapter 6 was a short primary research chapter aiming to answer **RQ. C** "*How can the spatial clustering of landslides in triggered event inventories be measured and is this 'generally applicable' behaviour?*". This research question was motivated by the visual disparity in the clustering of synthetic triggered landslide event inventories created in the LRIM model (**Chapter 7**) and real triggered landslide event inventories used to confront the model output. These observed visual disparities suggested that landslides cluster at a smaller spatial scale than captured by the LRIM; similar results have also been observed by Zaitchick *et al.* (2003). In terms of answering **RQ. C**, **Chapter 6** presents two methods (pair-correlation and kappa-measure) to quantify the spatial clustering in triggered landslide event inventories.

In the first part of **Chapter 6**, we showed how pair correlation (commonly used in a variety of disciplines) can be used to measure the distances (r) between the centroids of all landslides in an inventory, and applied this technique to eight triggered landslide event inventories, varying in location, triggering mechanism, number of landslides and confidence in the completeness of the inventory (as discussed in **Chapter 2**). For each

inventory examined, the probability density distribution of distances between landslide centroids ($p(r)$) differed from that of a homogeneously distributed set of randomly generated points. The probability density distribution of distances ($p(r)$) for randomly generated points was found to be symmetrical about the mean distance between points, whereas $p(r)$ for observed landslide centroids was found to be skewed towards smaller distances (maximum $p(r)$ ranged $1.2 < p(r) < 40.9$ km depending on the extent of the study region), suggesting that landslides cluster at a fine spatial scale. Where **Chapter 6** contributes to existing knowledge of landslide clustering is by showing that the shape of the probability density distribution of distances ($p(r)$) is quite similar between inventories. When normalised by the aerial extent over which the landslides were triggered, there appear to be some small variations in the pair-correlation distribution of landslides in low-mobility (inventories dominated by slides) and high-mobility (inventories dominated by rockfalls and debris flows) triggered landslide events. This complements the findings of **Chapter 5** that the statistical behaviour of landslide L/W and clustering may be 'semi generally-applicable' if separated into broad types of triggering mechanism.

The second part of **Chapter 6** developed a method (the Kappa-measure) to investigate the link between landslide susceptibility and number of landslides within a slope unit for the Collazzone (Italy) and Northridge (USA) triggered landslide event inventories. The kappa-measure works by placing a 1 km × 1 km grid over the study region and moving this by 200 m in the north direction, then the east direction, and finally in the north plus east direction, in a moving-window type approach. Within each grid cell, the proportion of the cell covered by the study area, the number of landslide centroids in that cell and the median landslide susceptibility (S_{LS}) in that cell are calculated. If < 0.25 (i.e., < 25 %) of the cell is covered by study area, the cell is rejected from further analysis. For cells where the proportion of the cell covered by study area is 0.25−1.00, the number of landslide centroids in that cell is normalised to the number of centroids per km^2 to give the kappa-measure of normalised number of landslide centroids per km^2 (κ_{LS}). By comparing κ_{LS} and median cell landslide susceptibility (S_{LS}), we showed that there is not a direct positive relationship between S_{LS} and landslide clustering, and the link between the two variables is noisy. For example, in some cells where $S_{LS} < 0.2$, $\kappa_{LS} = 5$−10 landslides km^{-2} and in other cells where $S_{LS} > 0.8$, $\kappa_{LS} = 0$ landslides km^{-2}. We believe this kappa measure contributes to existing understanding of the relationship between landslide susceptibility maps and triggered landslide events: that some high susceptibility regions will fail multiple times and some will not fail at all during a single triggered landslide event. The Kappa-measure may allow an

approximation of what proportion of high susceptibility landslide zones may fail during a single event, and the overall distribution of landslides per square kilometre within specific susceptibility classes. We hope that others will apply the kappa measure to locations where substantially complete triggered landslide event inventories and landslide susceptibility maps are available, to further test whether this is 'generally applicable' behaviour. The final section of **Chapter 6** showed some initial trials of incorporating clustering into the Landslide Road Impact Model (LRIM), showing that this introduction of clustering did not significantly improve model skill in forecasting number of road blockages resulting from a triggered landslide event.

8.2.3 Part III: Development of a Landslide-Road Impact Model

The final **Part III** of the thesis (**Chapter 7**) aims to answer **RQ. D** *"Is it possible to create a 'generally applicable' model of triggered landslide event impact upon regional road networks using the statistical behaviour of landslide size and shape?"*. We answered **RQ. D** by showing the development of a landslide-road impact model (LRIM) that creates 'synthetic' triggered landslide event inventory maps which reflect the generally observed statistical behaviour of triggered landslide events, and applying the LRIM to three study regions that vary in terms of geography, landslide triggering mechanism and economic status (Northridge, USA; Collazzone, Italy; Shu-Wa, Taiwan). Landslide areas and shapes are randomly selected from general statistical probability density functions and semi-randomly dropped over three regions that have been conditioned by landslide susceptibility. In the LRIM, the resultant synthetic triggered landslide event inventories are overlaid on to road network maps and impact identified. Model scenarios are run 100 times in a Monte Carlo simulation to create probabilistic forecasts of road network impact, which are confronted with observed triggered landslide events. By adjusting road network corridor buffers for local susceptibility, the model closely reflects the real-world number of road blockages. This landslide-road impact model presents an open-source low-data methodology for the stochastic simulation of potential simultaneous road network impacts.

We showed through confronting the LRIM model output with real observed data of road blockages from triggered landslide event inventories (in Collazzone, Central Italy; Oat Mountain, California USA; and Su-Hua, Taiwan) that there appears to be a complex relationship between road building and landslide occurrence. It is not new to find that road building may increase landslide incidence through slope modification and poor drainage (e.g., Brenning *et al.*, 2015). However, we do not believe that many others have discussed

the idea of road building inhibiting landslide occurrence in some locations. This may seem obvious to those more familiar with the engineering literature, as road protection and drainage measures may be integrated into road building (e.g., Hearn *et al.*, 2011). However, when fine scale maps of road engineering are either not available or kept by various bodies, it is difficult to incorporate this landslide-road interaction into landslide susceptibility maps. Thus, we believe the method we developed (as part of LRIM) to randomly buffer portions of the road network and reduce susceptibility within those zones may be a useful technique for others to approximate the effect of landslide road protection measures on landslide susceptibility.

Chapter 7 was the largest piece of research performed in this thesis, with the primary research on landslide shape and clustering in **Chapters 5** and **6** undertaken to inform the modelling process in **Chapter 7**. Consequently, we believe our more substantial contributions to knowledge are in this chapter. We outline these now:

1. *By developing a model to forecast landslide-road impact that requires minimal data inputs.* In **Section 8.1**, we reviewed Guzzetti *et al.* (1999)'s arguments that (i) quality data and (ii) appropriate methodologies to use analyse data are limiting factors in assessing landslide risk. Indeed, we showed in **Chapter 7** that much of the existing work on forecasting landslide-road impact requires large amounts of data, and/or is uncertain because of data issues. Through the semi-stochastic statistical approach used, the LRIM tool that we have developed effectively circumvents the issues of (i) lack of data or (ii) poor quality data. Moreover, because of the semi-stochastic Monte-Carlo simulation approach used in the LRIM, we are also able to give an estimate of uncertainty of model output, which is useful when dealing with uncertain data and thinking about different scenarios of impact.
2. *By demonstrating how the 'generally applicable' statistical behaviour of triggered landslide event inventories can be used in an applied sense.* The majority of papers reviewed in **Chapter 4** (statistical models of landslide area) focused on *why* we observed a general statistical behaviour for the probability of landslide areas in a substantially complete triggered landslide event, rather than what we can do with the general statistical distribution found. Certainly, the 'generally applicable' behaviour of landslide shape and clustering established in **Chapters 5** and **6** of this thesis was subject to uncertainty, and we were not able to fully explain the physical mechanism governing this behaviour. However, if we accept this generally applicable behaviour is in fact real,

we have demonstrated how this can be powerful for simulating triggered landslide event inventories and their impact. The LRIM model methodology could now be applied to other elements (e.g., buildings, other infrastructure, rivers) to forecast landslide risk.

3. *By uniting approaches from transportation and other hazard research*. The literature review in **Chapter 7** showed that other research communities (particularly for earthquakes) are much more advanced in considering multiple simultaneous failures resulting from one event, and also the network wide implications for the transportation infrastructure (e.g., Chang and Nojima, 2001). By borrowing techniques such as Monte-Carlo simulations and network analysis indices, we believe the LRIM forecast of triggered landslide event impact is more holistic because it gives scenarios of multiple road blockages from one triggered landslide event. We have also suggested ways of measuring the change in network structure following a triggered landslide event. This could be useful (i) before a triggered landslide event for thinking about building in extra redundancy to road networks to reduce the risk of isolation or (ii) during/after a triggered landslide event to provide a first estimation of impact and network accessibility.

We hope that others will now build upon and develop the LRIM methodology.

8.3 Concluding Remarks

Through the research and reviews performed in this thesis, we have shown that:

(i) *Landslide inventory (and other types of) data are open to many uncertainties*, and these uncertainties will affect the results of scientific and applied research.

(ii) *Analysis of landslide risk and impact often requires large amounts of detailed, locally specific physical and social data*, which is not often feasible to collect for many regions that are particularly vulnerable to triggered landslide events.

(iii) *We may address this shortfall* in landslide data to some extent through:

   a. *Robust searching of local newspaper articles* to find additional records of landslides and richer information about their impacts

   b. *The identification of 'generally applicable' behaviour*, developed from locations where we have large quantities of high quality data (such as large, substantially complete triggered landslide event inventories).

c. *The exploitation of 'generally applicable' statistical behaviour* of triggered landslide event inventories. This behaviour can be used as a 'template' to simulate triggered landslide event inventories in locations where high quality data does not exist, as shown through the development of the Landslide-Road Impact Model (LRIM).

These methods to address data shortfalls undeniably have their own uncertainties and biases, such as the following:

- Newspaper articles being focused on 'landslides with consequences' (Guzzetti *et al.*, 2003) discussed in **Chapter 2**.
- Difficulties in finding 'definitive' generally applicable statistical patterns in landslide shape without the analysis of additional large, substantially complete triggered landslide event inventories, discussed in **Chapter 5**.
- Disparities between LRIM model output and observed number of road blockages without factoring in measures of landslide road protection, discussed in **Chapter 7**.

However, throughout the primary research chapters of this thesis, we have shown the development of novel methods and tools (and tried to ensure their robustness) which can now be applied to other datasets. We have tried to ensure these methods are shared through the creation of open-source tools (and in time, manuals). With an appropriate consideration of the limitations (such as those listed above), users potentially have new approaches to examining and forecasting triggered landslide events. For locations where good data already exists (e.g., California), the methods presented here can be used to analyse existing datasets, and the LRIM may provide an interesting way to consider simultaneous failures on the transportation network, which have not been investigated previously. For locations where there is little access to data, the LRIM may provide a first estimate of landslide-road network impact, and a tool to visualise and discuss different scenarios of impact.

It is difficult to say with confidence whether we have established new 'generally applicable' patterns in triggered landslide event inventory shape and spatial clustering (**Chapters 5** and **6**). Indeed, the results appear to be less clearly 'generally applicable' than those in the literature for landslide area. We are unsure whether the 'noise' in the statistical patterns observed for landslide shape and clustering reveals that not all triggered landslide event

inventory behaviour is general, or that this simply reinforces the argument that analysing lower-confidence landslide inventories can result in more variable results.

The approaches in the research chapters of this thesis (particularly **Chapters 5, 6** and **7**) have very much been 'generalist' (i.e., favouring broad or simple approaches rather than detailed or complex investigations). We have at times actively avoided analysing potentially confounding additional datasets such as geology, land use, soil parameters, moisture etc. We made this decision based on the fact that there are very few places in the world (developed or developing) where there is the motivation to collect detailed, homogeneous observations of triggered landslide events and their impacts, so simple 'rule of thumb' solutions are more useful than globally optimizing, data hungry solutions. This novel and simple approach may explain why the Landslide-Road Impact Model (LRIM) has attracted attention from insurers, local governments, civil protection agencies and scientists over the year preceding submission of this PhD thesis. It is possible that through adopting a more physically-based, deterministic approach, others may be able to better explain *why* we observed some of the statistical behaviours that we did, such as the patterns in landslide shape, or the disparity between modelled and observed number of road blockages. We welcome others to use and adapt the tools and approaches that we have developed here to explore the physical processes resulting in this generally applicable statistical behaviour of landslides.

In summary, this thesis has attempted to document and share the development of generally applicable methodologies and tools (both in the text here and the open-source scripts created and included in the appendices) which in turn help to understand landslide risk when there are issues of data quality and quantity. The approaches used are inter-disciplinary: *digital humanities* techniques used in **Part I** to interrogate an archive of newspapers; *statistical and physics techniques* used in **Part II** to investigate the statistical behaviour of triggered landslide event inventories; and *civil-protection, spatial analysis, statistical modelling and network analysis techniques* used in **Part 3** to develop an applied model to understand landslide risk. By uniting a range of techniques from these varied disciplines, we have offered new ways of thinking about landslides that can be used globally to reduce their impact.

9. References Cited

Adams, D. C. and Rohlf, F. J. (2000) 'Ecological Character Displacement in Plethodon: Biomechanical Differences Found from a Geometric Morphometric Study' *Proceedings of the National Academy of Sciences,* 97**(8)**, p.4106-4111.

Adams, D. C., Rohlf, F. J., Slice, D. E. (2004) 'Geomorphic Morphometrics: ten years of progress following the "revolution"' *Italian Journal of Zoology*, 71**(1)**, p.5-16.

Akgun, A. (2012) 'A comparison of landslide susceptibility maps produced by logistic regression, multi-criteria decision, and likelihood ratio methods: a case study at İzmir, Turkey' *Landslides*, 9**(1)**, p.93-106.

Albert, R., Albert, I. and Nakarado, G. L. (2004) 'Structural Vulnerability of the North American Power Grid' *Physical Review E,* 69**(2)**, p.025103.

Alberto, C., Giovanni, C. and Paolo, F. (2003) 'Geomorphological and Historical Data in Assessing Landslide Hazard' *Earth Surface Processes and Landforms,* 28**(10)**, p.1125-1142.

Alcántara-Ayala, I. (2002) 'Geomorphology, Natural Hazards, Vulnerability and Prevention of Natural Disasters in Developing Countries' *Geomorphology,* 47**(2–4)**, p.107-124.

Aleotti, P.and Chowdhury, R. (1999) 'Landslide Hazard Assessment: Summary Review and New Perspectives' *Bulletin of Engineering Geology and the Environment,* 58**(1)**, p.21-44.

Allan, S., Adam, B. and Carter, C. (2013) 'Introduction: The Media Politics of Environmental Risk' In: Allan, S., Adam, B. and Carter, C. (eds.) *Environmental Risks and the Media.* London: Routledge. p.1-26.

Alvioli, M., Marchesini, I., Fiorucci, F., Rossi, M., Reichenbach, P.and Guzzetti, F. (2014) 'Automatic Delineation of Geo-Morphological Slope Units' (EGU2014-12697) Presented at the European Geosciences Union General Assembly 2014, 27 April-2 May, Vienna, Austria.

Amin, S. and Goldstein, M. P . (2008) *Data against Natural Disasters: Establishing Effective Systems for Relief, Recovery, and Reconstruction.* Washington DC: World Bank Publications.

Andrews, J. and Estabrook, G. (1971) 'Applications of Information and Graph Theory to Multivariate Geomorphological Analyses' *The Journal of Geology*, p.207-221.

Apel, H., Thieken, A., Merz, B. and Blöschl, G. (2006) 'A Probabilistic Modelling System for Assessing Flood Risks' *Natural hazards,* 38**(1-2)**, p.79-100.

Ardizzone, F., Cardinali, M., Carrara, A., Guzzetti, F. and Reichenbach, P . (1999) 'Impact of Mapping Errors on the Reliability of Landslide Hazard Maps' *Natural Hazards and Earth System Sciences*, 2**(1/2)**, p.3-14.

Ardizzone, F., Cardinali, M., Galli, M., Guzzetti, F. and Reichenbach, P . (2007) 'Identification and Mapping of Recent Rainfall-Induced Landslides Using Elevation Data Collected by Airborne LiDAR' *Natural Hazards and Earth System Sciences,* 7**(6)**, p.637-650.

Ardizzone, F., Cardinali, M., Guzzetti, F. and Reichenbach, P . (2008) 'Landslide Hazard Assessment, Vulnerability Estimation, and Risk Evaluation at the Basin Scale' In: *Proceedings of the First World Landslide Forum, Parallel Session*, vol. 2008, pp.71-74. 2008.

Ardizzone, F., Fiorucci, F., Reichenbach, P., Mateos, R. M. (2014b) 'Impact Assessment of Landslide Events on Structures and Infrastructures' In: Malamud, B. D., Mota, B., Aridzzone, F., Taylor, F. E., Drake, N., Reichenbach, P., Fiorucci, F., Mateos, R. M. (2014) *LAMPRE Deliverable 6.4 Report on Methods for Different Sized Events and Different Type Road Networks* [online] Available at: http://www.lampre-project.eu/index.php?option=com_phocadownload&view=category&id=27:wp6-preparedeness-prevention-recovery-reconstruction&Itemid=525 [Accessed 27/04/2015]., p.22-44.

Ardizzone, F., Rossi, M., Cardinali, M., Marchesini, I., Santangelo, M., Mondini, A., Fiorucci, F . (2014a) 'Standards for Landslide Inventory Maps and Quality Assessment' Report D5.1 LAMPRE EU FP7 Project [online] Available at: http://www.lampre-project.eu/index.php?option=com_phocadownload&view=category&id=26:wp5-triggered-event-landslides&Itemid=525 [Accessed 27/04/2015].

Ash, J. and Newth, D. (2007) 'Optimizing Complex Networks for Resilience against Cascading Failure' *Physica A: Statistical Mechanics and its Applications,* 380**(0)**, p.673-683.

Ayalew, L., and Yamagishi, H. (2005) 'The application of GIS-based logistic regression for landslide susceptibility mapping in the Kakuda-Yahiko Mountains, Central Japan' *Geomorphology*, 65**(1)**, p.15-31.

Bak, P., Tang, C. and Wiesenfeld, K. (1987) 'Self-Organized Criticality: An Explanation of the 1/F Noise' *Physical Review Letters,* 59**(4)**, p.381-384.

Barlow, J. Franklin, S., Martin, Y. (2006) 'High Spatial Resolution Satellite Imagery, DEM Derivatives, and Image Segmentation for the Detection of Mass Wasting Processes' *Photogrammetric Engineering and Remote Sensing*, 72**(6)**, p.687-692.

Barlow, J., Lim, M., Rosser, N., Petley, D., Brain, M., Norman, E. and Geer, M. (2012) 'Modelling Cliff Erosion Using Negative Power Law Scaling of Rockfalls' *Geomorphology,* 139–140**(0)**, p.416-424.

Baroni, C., Barabegoli, E., Onorevoli, G. (2015) 'Temporal Reliability of a Landslide Inventory Map. A Case Study in the Northern Apennines (Emilia-Romagna Region, Italy)' (EGU2015-2645) Presented at the European Geosciences Union General Assembly 2015, 12-17 April, Vienna, Austria.

Barredo, J. (2007) 'Major Flood Disasters in Europe: 1950–2005' *Natural Hazards,* 42**(1)**, p.125-148.

Barton, M. E., Coles, B. J. and Tiller, G. R. (1983) 'A Statistical Study of the Cliff Top Slumps in Part of the Christchurch Bay Coastal Cliffs' *Earth Surface Processes and Landforms,* 8**(5)**, p.409-422.

Baxter, J. and Eyles, J. (1997) 'Evaluating Qualitative Research in Social Geography: Establishing 'Rigour' in Interview Analysis' *Transactions of the Institute of British Geographers,* 22**(4)**, p.505-525.

Bayliss, A. C. and Jones, R. C. (1993) *Peaks-over-Threshold Flood Database: Summary Statistics and Seasonality*. IH Report No. 121, Wallingford: Institute of Hydrology.

Bear, J. (2013) *Dynamics of Fluids in Porous Media*. New York: American Elsevier Publishing Company Inc.

Bednarik, M., Magulová, B., Matys, M. and Marschalko, M. (2010) 'Landslide Susceptibility Assessment of the Kraľovany–Liptovský Mikuláš Railway Case Study' *Physics and Chemistry of the Earth, Parts A/B/C,* 35**(3)**, p.162-171.

Bell, R., Petschko, H., Roehrs, M. and Dix, A. (2012) 'Assessment of Landslide Age, Landslide Persistence and Human Impact Using Airborne Laser Scanning Digital Terrain Models' *Geografiska Annaler: Series A, Physical Geography,* 94**(1)**, p.135-156.

Benediktsson, J. A., Pesaresi, M. and Amason, K. (2003) 'Classification and Feature Extraction for Remote Sensing Images from Urban Areas Based on Morphological Transformations' *Geoscience and Remote Sensing, IEEE Transactions on,* 41**(9)**, p.1940-1949.

BGS (2014a) 'BGS National Landslide Database' *British Geological Survey* [online] Available at: http://www.bgs.ac.uk/landslides/nld.html [Accessed 23/04/2015].

BGS (2014c) 'National Landslide Potential Map (Geosure)' *British Geological Survey* [online] Available at: http://www.bgs.ac.uk/products/geosure/ [Accessed 23/04/2015].

BGS (2015) 'Terrestrial LiDAR image of Holderness to Spurn Head coast' [online] Available at: http://www.bgs.ac.uk/research/climatechange/environment/coastal/coastal Erosion.html [Accessed 27/04/2015].

Black, A. R. and Law, F. M. (2004) 'Development and Utilization of a National Web-Based Chronology of Hydrological Events' *Hydrological Sciences Journal-Journal Des Sciences Hydrologiques,* 49**(2)**, p.237-246.

Blackford, J. J. and Chambers, F. M. (1991) 'Proxy Records of Climate from Blanket Mires: Evidence for a Dark Age (1400 Bp) Climatic Deterioration in the British Isles' *The Holocene,* 1**(1)**, p.63-67.

Blasio, F. V. (2011) *Introduction to the Physics of Landslides*, New York: Springer

Bollobás, B. e., Riordan, O., Spencer, J. and Tusnády, G. (2001) 'The Degree Sequence of a Scale-Free Random Graph Process' *Random Structures & Algorithms,* 18**(3)**, p.279-290.

BootAnim (2015) 'Bootstrap Animations' [online] Available at: https://www.stat.auckland .ac.nz/ ~wild/BootAnim/ [Accessed 27/04/2015].

Borghuis, A. M., Chang, K. and Lee, H. Y. (2007) 'Comparison between Automated and Manual Mapping of Typhoon-Triggered Landslides from Spot-5 Imagery' *International Journal of Remote Sensing,* 28**(8)**, p.1843-1856.

Borgomeo, E., Hebditch, K. V., Whittaker, A. C. and Lonergan, L. (2014) 'Characterising the Spatial Distribution, Frequency and Geomorphic Controls on Landslide Occurrence, Molise, Italy' *Geomorphology,* 226**(0)**, p.148-161.

Boruff, B. J. (2009) 'Environmental Hazards: Assessing Risk and Reducing Disasters, 5th Edition – by Keith Smith and David N. Petley' *Geographical Research,* 47**(4)**, p.454-455.

Brardinoni, F. and Church, M. (2004) 'Representing the Landslide Magnitude–Frequency Relation: Capilano River Basin, British Columbia' *Earth Surface Processes and Landforms,* 29**(1)**, p.115-124.

Brardinoni, F., Slaymaker, O. and Hassan, M. A. (2003) 'Landslide Inventory in a Rugged Forested Watershed: A Comparison between Air-Photo and Field Survey Data' *Geomorphology,* 54**(3–4)**, p.179-196.

Brázdil, R., Pfister, C., Wanner, H., Von Storch, H. and Luterbacher, J. (2005) 'Historical climatology in Europe–the state of the art' *Climatic change*, 70**(3),** p.363-430.

Brenning, A., Schwinn, M., Ruiz-Páez, A. P., and Muenchow, J. (2015) 'Landslide susceptibility near highways is increased by 1 order of magnitude in the Andes of southern Ecuador, Loja province' *Natural Hazards and Earth System Science*, 15**(1)**, p.45-57.

British Library. (2014) 'The British Newspaper Archive' [online] Available at: http://www.british newspaperarchive.co.uk/ [Accessed 27/04/2015].

Bromhead, E. N. and Ibsen, M. L. (2006) 'A Review of Landsliding and Coastal Erosion Damage to Historic Fortifications in South East England' *Landslides,* 3**(4)**, p.341-347.

Brunetti, M. T., Guzzetti, F., Cardinali, M., Fiorucci, F., Santangelo, M., Mancinelli, P., Komatsu, G. and Borselli, L. (2014) 'Analysis of a New Geomorphological Inventory of Landslides in Valles Marineris, Mars' *Earth and Planetary Science Letters,* 405, p.156-168.

Brunetti, M., Guzzetti, F. and Rossi, M. (2009) 'Probability Distributions of Landslide Volumes' *Nonlinear Processes in Geophysics,* 16**(2)**, p.179-188.

Brunsden, D. 1985. Landslide Types, Mechanisms, Recognition, Identification. In: Morgan (ed.) *Landslides in the South Wales Coalfield : Proceedings, Symposium, Polytechnic of Wales, 1st to 3rd April 1985*. Pontypridd: Polytechnic of Wales, 1985 (1986 printing) p.19–28.

Bucknam, R. C., Coe, J. A., Chavarría, M. M., Godt, J. W., Tarr, A. C., Bradley, L., Rafferty, S., Hancock, D., Dart, R. and Johnson, M. (2001) *Landslides Triggered by Hurricane Mitch in Guatemala: Inventory and Discussion*. Open File Report 01-443, US Department of the Interior, US Geological Survey.

Butler, D. (2013) 'Nature News: Crowdsourcing Goes Mainstream in Typhoon Response' [online] Available at: http://www.nature.com/news/crowdsourcing-goes-mainstream-in-typhoon-response-1.14186 [Accessed 27/04/2015].

Cardinali, M., Antonini, G., Reichenbach, P.and Guzzetti, F. (2001) 'Photo-Geological and Landslide Inventory Map for the Upper Tiber River Basin' *CNR, Gruppo Nazionale per la Difesa dalle Catastrofi Idrogeologiche,* Publication 2154.

Cardinali, M., Ardizzone, F., Galli, M., Guzzetti, F. and Reichenbach, P . (2000) 'Landslides Triggered by Rapid Snow Melting: The December 1996–January 1997 Event in Central Italy' In: Claps, P.and Siccardi, F. (eds.) *Proceedings 1st Plinius Conference on Mediterranean Storms,* Bios: Cosenza, p.439–448.

Cardinali, M., Galli, M., Guzzetti, F., Ardizzone, F., Reichenbach, P.and Bartoccini, P . (2006) 'Rainfall Induced Landslides in December 2004 in South-Western Umbria, Central Italy: Types, Extent, Damage and Risk Assessment' *Natural Hazards and Earth System Sciences,* 6**(2)**, p.237-260.

Carrara, A., Cardinali, M., Detti, R., Guzzetti, F., Pasqui, V. and Reichenbach, P . (1991) 'GIS Techniques and Statistical Models in Evaluating Landslide Hazard' *Earth Surface Processes and Landforms,* 16**(5)**, p.427-445.

Carrara, A., Cardinali, M., Guzzetti, F. and Reichenbach, P . (1995) 'GIS Technology in Mapping Landslide Hazard' In: Carrara, A. and Guzzetti, F (eds.) *Geographical Information Systems in Assessing Natural Hazards*. Netherlands: Springer, p.135-175.

Carrara, A., Crosta, G. and Frattini, P . (2003) 'Geomorphological and Historical Data in Assessing Landslide Hazard' *Earth Surface Processes and Landforms,* 28**(10)**, p.1125-1142.

Carvalho, A. (2007) 'Ideological Cultures and Media Discourses on Scientific Knowledge: Re-Reading News on Climate Change' *Public Understanding of Science,* 16**(2)**, p.223-243.

Casadei, M., Dietrich, W. and Miller, N. (2003) 'Controls on Shallow Landslide Size' *In: Proceedings of the 3rd International Conference on Debris-Flow Hazards Mitigation: Mechanics, Prediction, and Assessment, Davos, Swizerland*, 91-101.

Catani, F., Casagli, N., Ermini, L., Righini, G. and Menduni, G. (2005) 'Landslide Hazard and Risk Mapping at Catchment Scale in the Arno River Basin' *Landslides,* 2**(4)**, p.329-342.

Cats, O. and Jenelius, E. (2015) 'Planning for the Unexpected: The Value of Reserve Capacity for Public Transport Network Robustness' *Transportation Research Part A: Policy and Practice*.

Cevik, E. and Topal, T. (2003) 'GIS-Based Landslide Susceptibility Mapping for a Problematic Segment of the Natural Gas Pipeline, Hendek (Turkey)' *Environmental Geology,* 44**(8)**, p.949-962.

Cha, E. J. and Ellingwood, B. R. (2012) 'Risk-Averse Decision-Making for Civil Infrastructure Exposed to Low-Probability, High-Consequence Events' *Reliability Engineering & System Safety,* 104**(0)**, p.27-35.

Cha, S. H. (2007) 'Comprehensive Survey on Distance/Similarity Measures between Probability Density Functions' *City,* 1**(2)**, p.1.

Chakraborty, I., Ghosh, S., Debasish, B. and Bora, A. (2011) 'Earthquake Induced Landslides in the Sikkim-Darjeeling Himalayas' [online] Available at: http://earthquake-report.com/2011/11/25/earthquake-induced-landslides-in-the-sikkim-darjeeling-himalayas/ [Accessed 27/02/2015].

Chang, S. E. and Nojima, N. (2001) 'Measuring Post-Disaster Transportation System Performance: The 1995 Kobe Earthquake in Comparative Perspective' *Transportation Research Part A: Policy and Practice,* 35**(6)**, p.475-494.

Chen, C.-Y. (2009) 'Sedimentary Impacts from Landslides in the Tachia River Basin, Taiwan' *Geomorphology,* 105**(3–4)**, p.355-365.

Chen, X.-L., Liu, C.-G., Chang, Z.-F. and Zhou, Q. (2015) 'The Relationship between the Slope Angle and the Landslide Size Derived from Limit Equilibrium Simulations' *Geomorphology,* Advance online publication, doi:10.1016/j.geomorph.2015.01.036.

Cheng, L., English, M. and Wong, R. (1985) 'Hailstone Size Distributions and Their Relationship to Storm Thermodynamics' *Journal of Climate and Applied Meteorology,* 24**(10)**, p.1059-1067.

Christopoulos, A. and Lew, M. J. (2000) 'Beyond Eyeballing: Fitting Models to Experimental Data' *Critical Reviews in Biochemistry and Molecular Biology,* 35**(5)**, p.359-391.

Chung, C. J. F., and Fabbri, A. G. (1999) 'Probabilistic prediction models for landslide hazard mapping' *Photogrammetric engineering and remote sensing*, 65**(12)**, p.1389-1399.

CISIS. (2014) 'Dbprior10k Map Portal' [online] Available at: http://www.centrointerregionale-gis.it/DBPrior/DBPrior.asp [Accessed 27/04/2015].

Clauset, A., Shalizi, C. R. and Newman, M. E. (2009) 'Power-Law Distributions in Empirical Data' *SIAM review,* 51**(4)**, p.661-703.

Clayton, C., Simons, N., Matthews, M. (1982) *Site Investigation*, London: Granada Publishing Ltd.

CNR-GNDCI (2012) 'National Map Portal for Hydrogeological Disasters' [online] Available at: http://www.gndci.cnr.it/ [Accessed 27/04/2015].

Coe, J., Godt, J., Baum, R., Bucknam, R. and Michael, J. (2004) 'Landslide Susceptibility from Topography in Guatemala' *Landslides: evaluation and stabilization,* 1**,** p.69-78.

Collins, B. D., Kayen, R. and Tanaka, Y. (2012) 'Spatial Distribution of Landslides Triggered from the 2007 Niigata Chuetsu–Oki Japan Earthquake' *Engineering Geology,* 127**(0)**, p.14-26.

Corominas, J., Remondo, J., Farias, P., Estevao, M., Zezere, J., Diaz de Teran, J., Dikau, R., Schrott, L., Moya, J. and Gonzalez, A. (1996) 'Debris Flow' In Dikau, R., Brunsden, D., Schrott L. and Ibsen, M. (eds.) *Landslide Recognition, Identification, Movement and causes.* New York: Wiley and sons. p.161.

Cova, T., Conger, S. (2004) 'Transportation Hazards' in Kutz, M. (ed.) *Handbook of Transportation Engineering*, New York: McGraw-Hill, Chapter 17.

Cover, T. and Hart, P . (1967) 'Nearest Neighbor Pattern Classification' *Information Theory, IEEE Transactions on,* 13**(1)**, p.21-27.

Crone, A. J., Baum, R. L., Lidke, D. J., Sather, D. N. D., Bradley, L-A., Tarr, A. C. (2001)' Landslides Induced by Hurricane Mitch in El Salvador – An Inventory and Decriptions of Selected Features' Open File Report 01-444, *Denver: United States Geological Survey.*

Crosta, G. B., Dal Negro, P.and Frattini, P . (2003) 'Soil Slips and Debris Flows on Terraced Slopes' *Natural Hazards and Earth System Sciences*, 3, p.31-42.

Crosta, G., Chen, H. and Lee, C. (2004) 'Replay of the 1987 Val Pola Landslide, Italian Alps' *Geomorphology,* 60**(1)**, p.127-146.

Crovelli, R. A., and Coe, J. A. (2009) 'Probabilistic estimation of numbers and costs of future landslides in the San Francisco Bay region' *Georisk*, 3**(4)**, p.206-223.

Crozier, M. J. (2005) 'Multiple-Occurrence Regional Landslide Events in New Zealand: Hazard Management Issues' *Landslides,* 2**(4)**, p.247-256.

Cruden, D. and Varnes, D. J. (1996) 'Landslide Types and Processes' In: Turner, K. A., and Schuster, R. L. (eds.) *Landslides: Investigation and Mitigation.* Washington D.C.: Transportation Research Board. p.36-75.

Crutcher, H. L. (1975) 'A Note on the Possible Misuse of the Kolmogorov-Smirnov Test' *Journal of Applied Meteorology,* 14**(8)**, p.1600-1603.

Czuchlewski, K. R., Weissel, J. K. and Kim, Y. (2003) 'Polarimetric Synthetic Aperture Radar Study of the Tsaoling Landslide Generated by the 1999 Chi-Chi Earthquake, Taiwan' *J. Geophys. Res.,* 108**(F1)**, p.6006.

Dai, F. C. and Lee, C. F. (2001) 'Frequency–Volume Relation and Prediction of Rainfall-Induced Landslides' *Engineering Geology,* 59**(3–4)**, p.253-266.

Dai, F. C., Lee, C. F. and Ngai, Y. Y. (2002) 'Landslide Risk Assessment and Management: An Overview' *Engineering Geology,* 64**(1)**, p.65-87.

Dai, F. C., Xu, C., Yao, X., Xu, L., Tu, X. B. and Gong, Q. M. (2011) 'Spatial Distribution of Landslides Triggered by the 2008 Ms 8.0 Wenchuan Earthquake, China' *Journal of Asian Earth Sciences,* 40**(4)**, p.883-895.

Damm, B., and Klose, M. (2015) 'The landslide database for Germany: Closing the gap at national level' *Geomorphology*, Advance Article in Press, doi:10.1016/j.geomorph.2015 .03.021.

Danko, D. M. (1992) 'The Digital Chart of the World Project' *Photogrammetric engineering and remote sensing,* 58**(8)**, p.1125-1128.

Danneels, G., Pirard, E. and Havenith, H. B. (2007) 'Automatic Landslide Detection from Remote Sensing Images Using Supervised Classification Methods' *In: Geoscience and Remote Sensing Symposium, 2007. IGARSS 2007. IEEE International*, 3014-3017.

Das, I. C. (2011) *Spatial Statistical Modelling for Assessing Landslide Hazard and Vulnerability*, Ph.D. Thesis, University of Twente.

Das, I., Stein, A., Kerle, N., and Dadhwal, V. K. (2012) 'Landslide susceptibility mapping along road corridors in the Indian Himalayas using Bayesian logistic regression models' *Geomorphology*, 179, p.116-125.

de Berg, M., Cheong, O., van Kreveld, M. and Overmars, M. (2008) *Computational Geometry: Algorithms and Applications*. Berlin: Springer.

De Groeve, T., Poljansek, K., and Ehrlich, D. (2013) *Recording Disaster Losses. Recommendations for a European Approach.* European Commission Joint Research Centre, Ispra: Italy.

De Leeuw, J., Said, M., Ortegah, L., Nagda, S., Georgiadou, Y. and DeBlois, M. (2011) 'An Assessment of the Accuracy of Volunteered Road Map Production in Western Kenya' *Remote Sensing,* 3**(2)**, p.247-256.

Deacon, D. (2007) 'Yesterday's papers and today's technology digital newspaper archives and 'push button' content analysis' *European Journal of Communication*, 22**(1)**, p.5-25.

DeGraff, V., Bryce, R., Jibson, R. Mora, S., Rogers, C. (1989) 'Landslides: Their Extent and Significance in the Caribbean' *In:* Brabb, E. and Harrold, B. (eds.) *Landslides: Their Extent and Ecological Significance*, Rotterdam: Balkerma, p.51-80.

Demeritt, D. (1991) 'Climate, cropping, and society in Vermont, 1820-1850' *Vermont History*, 50**(3)**, p.133-65.

Densmore, A. L. and Hovius, N. (2000) 'Topographic Fingerprints of Bedrock Landslides' *Geology,* 28**(4)**, p.371-374.

Densmore, A. L., Ellis, M. A., and Anderson, R. S. (1998) 'Landsliding and the evolution of normal-fault-bounded mountains' *Journal of Geophysical Research: Solid Earth*, *103***(B7)**, p.15203-15219.

Devoli, G., Morales, A. and Høeg, K. (2007) 'Historical Landslides in Nicaragua—Collection and Analysis of Data' *Landslides,* 4**(1)**, p.5-18.

Dewitte, O., Chung, C.-J. and Demoulin, A. (2006) 'Reactivation Hazard Mapping for Ancient Landslides in West Belgium' *Natural Hazards and Earth System Science,* 6**(4)**, p.653-662.

Dickinson, J. L., Zuckerberg, B., and Bonter, D. N. (2010) 'Citizen science as an ecological research tool: challenges and benefits' *Annual review of ecology, evolution, and systematics*, 41, p.149-172.

DiDonato, A. R. and Alfred H Morris, J. (1986) 'Computation of the Incomplete Gamma Function Ratios and Their Inverse' *ACM Trans. Math. Softw.,* 12**(4)**, p.377-393.

Dikau, R. (1996) *Landslide Recognition: Identification, Movement and Courses*. Chichester: Wiley.

Dikau, R., Bunsden, D., Schrott, L., Ibsen, M. (1996) 'Introduction' In: Dikau, R., Bunsden, D., Schrott, L., Ibsen, M. (eds.) *Landslide Recognition: Identification, Movement and Causes*, Chichester: John Wiley and Sons, p.1-13.

Dikau, R., Cavallin, A. and Jager, S. (1996) 'Databases and GIS for Landslide Research in Europe' *Geomorphology,* 15**(3-4)**, p.227-239.

Domínguez Cuesta, M. a. J., Jiménez Sánchez, M. and Rodríguez García, A. (1999) 'Press Archives as Temporal Records of Landslides in the North of Spain: Relationships between Rainfall and Instability Slope Events' *Geomorphology,* 30**(1–2)**, p.125-132.

Dueñas-Osorio, L., Craig, J. I. and Goodno, B. J. (2007) 'Seismic Response of Critical Interdependent Networks' *Earthquake Engineering & Structural Dynamics,* 36**(2)**, p.285-306.

Duman, T. Y., Çan, T., Emre, Ö., Keçer, M., Doğan, A., Ateş, Ş. and Durmaz, S. (2005) 'Landslide Inventory of Northwestern Anatolia, Turkey' *Engineering Geology,* 77**(1–2)**, p.99-114.

Efron, B. and Tibshirani, R. J. (1993) *An Introduction to the Bootstrap*. New York ; London: Chapman & Hall.

Ekström, G., and Stark, C. P . (2013) 'Simple scaling of catastrophic landslide dynamics' *Science*, 339**(6126),** p.1416-1419.

Elliott, A. and Kirschbaum, M. J. (2007) 'The Preliminary Landslide History Database of Utah, 1850-1978' Open File Report 514, Salt Lake City: Utah Geological Survey.

Escobar, M. P. and Demeritt, D. (2014) 'Flooding and the Framing of Risk in British Broadsheets, 1985–2010' *Public Understanding of Science*, p.454-471.

Esmen, N. A. and Hammad, Y. Y. (1977) 'Log-Normality of Environmental Sampling Data' *Journal of Environmental Science and Health . Part A: Environmental Science and Engineering,* 12**(1-2)**, p.29-41.

ESRI, USDA, USGS, AEX, GeoEye, Getmapping, AeroGRID, IGN, IGP, UPR-EGP; (2013) 'ESRI Basemaps and Reference Layers' [online] Available through ESRI Arc Map Base

Map Layers. Info at: http://www.esri.com/software/arcgis/arcgisonline/maps/maps-and-map-layers [Accessed 27/04/2015].

Fan, S., Chan-Kang, C. (2005) 'Road Development, Economic Growth and Poverty Reduction in China' Research Report 138, Washington DC: International Food Policy Research Institute.

Fan, X., van Westen, C. J., Korup, O., Gorum, T., Xu, Q., Dai, F., Huang, R. and Wang, G. (2012) 'Transient Water and Sediment Storage of the Decaying Landslide Dams Induced by the 2008 Wenchuan Earthquake, China' *Geomorphology,* 171–172**(0)**, p.58-68.

Farina, P., Colombo, D., Fumagalli, A., Marks, F. and Moretti, S. (2006) 'Permanent Scatterers for Landslide Investigations: Outcomes from the Esa-Slam Project' *Engineering Geology,* 88**(3–4)**, p.200-217.

Fell, R., Corominas, J., Bonnard, C., Cascini, L., Leroi, E., Savage, W. Z. and Eng, J.-J. T. C. L. (2008) 'Guidelines for Landslide Susceptibility, Hazard and Risk Zoning for Land-Use Planning Commentary' *Engineering Geology,* 102**(3-4)**, p.99-111.

FEMA (2003) MR3 Technical Manual. *Multi-hazard Loss Estimation Methodology Earthquake Model* Washington DC: United States Department of Homeland Security.

Fernández, T., Delgado, J., Cardenal, J., Irigaray, C., El Hamdouni, R., Chacón, J. (2006) 'Improvement of Positional Accuracy of a Landslide Database using Digital Photogrammetry Techniques' *In* Caetano, M. and Painho, M. (eds) *Proceedings of the 7th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences*, 5 – 7 July 2006, Lisboa, Instituto Geográfico Português, p.139-149.

Fiedrich, F., Gehbauer, F. and Rickers, U. (2000) 'Optimized Resource Allocation for Emergency Response after Earthquake Disasters' *Safety Science,* 35**(1–3)**, p.41-57.

Filipello, A., Strigaro, D. (2014) 'GRASS-GIS r.shalstab tool' [online] Available at: http://grass.osgeo.org/grass70/manuals/addons/r.shalstab.html [Accessed 27/04/2015].

Fiorucci, F., Cardinali, M., Carlà, R., Rossi, M., Mondini, A. C., Santurri, L., Ardizzone, F. and Guzzetti, F. (2011) 'Seasonal Landslide Mapping and Estimation of Landslide Mobilization Rates Using Aerial and Satellite Images' *Geomorphology,* 129**(1–2)**, p.59-70.

Fischer, H. W. (1994) *Response to Disaster: Fact Versus Fiction and Its Perpetuation.* Lanham: University Press of America.

Flageollet, J. C. (1996) 'Temporal Occurrence and Forecasting of Landslides in the European Community' *Hydrological and Hydrogeological Risks*, p.73-80.

Fonstad, M. A., Dietrich, J. T., Courville, B. C., Jensen, J. L. and Carbonneau, P.E. (2013) 'Topographic Structure from Motion: A New Development in Photogrammetric Measurement' *Earth Surface Processes and Landforms,* 38**(4)**, p.421-430.

Fookes, P., Sweeney, M., Manby, C. t. and Martin, R. (1985) 'Geological and Geotechnical Engineering Aspects of Low-Cost Roads in Mountainous Terrain' *Engineering Geology,* 21**(1)**, p.1-152.

Forman, R. T. T. (1995) *Land Mosaics : The Ecology of Landscapes and Regions*. Cambridge ; New York: Cambridge University Press.

Forman, R. T. T. and Alexander, L. E. (1998) 'Roads and Their Major Ecological Effects' *Annual Review of Ecology and Systematics,* 29, p.207-C2.

Foster, C., Gibson, A., Wildman, G. and British Geological Survey. (2008) 'The New National Landslide Database and Landslide Hazard Assessment of Great Britain' *In: First World Landslide Forum*, 203-206. Tokyo, Japan: ISDR.

Foster, C., Pennington, C. V. L., Culshaw, M. G. and Lawrie, K. (2012) 'The National Landslide Database of Great Britain : Development, Evolution and Applications' *Environmental Earth Sciences,* 66**(3)**, p.941-953.

Frattini, P.and Crosta, G. B. (2013) 'The Role of Material Properties and Landscape Morphology on Landslide Size Distributions' *Earth and Planetary Science Letters,* 361, p.310-319.

Freudenburg, W. R., Coleman, C.-L., Gonzales, J. and Helgeland, C. (1996) 'Media Coverage of Hazard Events: Analyzing the Assumptions' *Risk Analysis,* 16**(1)**, p.31-42.

Frigg, R. (2003) 'Self-Organised Criticality—What It Is and What It Isn't' *Studies In History and Philosophy of Science Part A,* 34**(3)**, p.613-632.

Fuji, Y. (1969) 'Frequency Distribution of Landslides Caused by Heavy Rainfall' *Journal of the Seismological Society of Japan*, 22, p.244-247.

FWHA (2011) 'Roadway Extent, Characteristics and Performance' US Department of Transportation Federal Highway Administration [online] Available at: http://www.fhwa.dot.gov/policy/ohim/hs05/htm/vm1.htm [Accessed 27/04/2015].

Gabet, E. J. and Dunne, T. (2002) 'Landslides on Coastal Sage-Scrub and Grassland Hillslopes in a Severe El Nino Winter: The Effects of Vegetation Conversion on Sediment Delivery' *Geological Society of America Bulletin,* 114**(8)**, p.983-990.

Galli, M. and Guzzetti, F. (2007) 'Landslide Vulnerability Criteria: A Case Study from Umbria, Central Italy' *Environmental Management,* 40**(4)**, p.649-665.

Galli, M., Ardizzone, F., Cardinali, M., Guzzetti, F. and Reichenbach, P . (2008) 'Comparing Landslide Inventory Maps' *Geomorphology,* 94**(3)**, p.268-289.

Gao, J. and Maro, J. (2010) 'Topographic Controls on Evolution of Shallow Landslides in Pastoral Wairarapa, New Zealand, 1979–2003' *Geomorphology,* 114**(3)**, p.373-381.

Garver, J. W. (2006) 'Development of China's Overland Transportation Links with Central, South-West and South Asia' *The China Quarterly,* 185, p.1-22.

GDACS. (2014) 'Global Disaster Alert and Coordination System' United Nations and the European Commission [online] Available at: http://www.gdacs.org/ [Accessed 27/04/2015].

Geobrugg (2012) 'TECCO® Stabilizes Slopes Susceptible to Weathering, Prevents Breakouts and Grows into One with Nature' [online] Available at: http://www.geobrugg.com /contento/English/Home/Slopestabilization/tabid/2061/language/en-US/Default.aspx [Accessed 06/03/2012].

Gerscovich, D. M. S and Costa, H. (2004) 'Drainage and Surface Protection' In: Ortigão, J.A. R., Sayao, A. S. F. J. (eds.) *Handbook of Slope Stabilisation*, Berlin: Springer-Verlag, p.147-211.

Ghil, M., Yiou, P., Hallegatte, S., Malamud, B. D., Naveau, P., Soloviev, A., Friederichs, P., Keilis-Borok, V., Kondrashov, D., Kossobokov, V., Mestre, O., Nicolis, C., Rust, H. W., Shebalin, P., Vrac, M., Witt, A. and Zaliapin, I. (2011) 'Extreme Events: Dynamics, Statistics and Prediction' *Nonlin. Processes Geophys.,* 18**(3)**, p.295-350.

Gibson, A. D., Culshaw, M. G., Dashwood, C. and Pennington, C. V. L. (2013) 'Landslide Management in the UK—the Problem of Managing Hazards in a 'Low-Risk' Environment' *Landslides,* 10**(5)**, p.599-610.

Glade, T. 2001. Landslide Hazard Assessment and Historical Landslide Data—an Inseparable Couple? *In, The Use of Historical Data in Natural Hazard Assessments*. Springer. p.153-168

Glade, T. and Crozier, M. (1996) 'Towards a National Landslide Information Base for New Zealand' *New Zealand Geographer,* 52**(1)**, p.29-40.

Glade, T., Crozier, M. J. (2005) 'The Nature of Landslide Hazard Impact' in Glade, T., Anderson, M., Crozier, M. J. (eds.) (2005) *Landslide Hazard and Risk*, Chichester: Wiley, Chapter 2, p.43-74.

GLCF (2006) 'Shuttle Radar Topography Mission 1 Arc-Second Elevation Map for Los Angeles Area, USA' Global Land Cover Facility, University of Maryland.

Glenn, N. F., Streutker, D. R., Chadwick, D. J., Thackray, G. D. and Dorsch, S. J. (2006) 'Analysis of LiDAR-Derived Topographic Information for Characterizing and Differentiating Landslide Morphology and Activity' *Geomorphology,* 73**(1-2)**, p.131-148.

GN (2014) 'Geoportale Nazionale Italian Map Portal' [online] Available at: http://www.pcn.minambiente.it/GN/ [Accessed 27/04/2015].

Goltz, C. (1996) 'Multifractal and Entropic Properties of Landslides in Japan' *Geologische Rundschau,* 85**(1)**, p.71-84.

Google Earth (2014) 'Various Imagery' [online] Available at: http://www.google.co.uk/intl/en_uk/earth/index.html [Accessed 27/04/2014].

Gordon, P., Richardson, H. W. and Davis, B. (1998) 'Transport-Related Impacts of the Northridge Earthquake' *Journal of Transportation and Statistics*, 1**(2)**, p.21-36.

Gorsevski, P.V., Gessler, P.E., Foltz, R. B. and Elliot, W. J. (2006) 'Spatial Prediction of Landslide Hazard Using Logistic Regression and Roc Analysis' *Transactions in GIS,* 10**(3)**, p.395-415.

Gorum, T., Fan, X., van Westen, C. J., Huang, R. Q., Xu, Q., Tang, C., and Wang, G. (2011) 'Distribution pattern of earthquake-induced landslides triggered by the 12 May 2008 Wenchuan earthquake' *Geomorphology*, 133**(3)**, 152-167.

Gorum, T., van Westen, C. J., Korup, O., van der Meijde, M., Fan, X. and van der Meer, F. D. (2013) 'Complex Rupture Mechanism and Topography Control Symmetry of Mass-Wasting Pattern, 2010 Haiti Earthquake' *Geomorphology,* 184**(0)**, p.127-138.

Greenberg, D. S., Houweling, A. R. and Kerr, J. N. D. (2008) 'Population Imaging of Ongoing Neuronal Activity in the Visual Cortex of Awake Rats' *Nat Neurosci,* 11**(7)**, p.749-751.

Greer, J., D. and Mensing, D. (2006) The Evolution of Online Newspapers: A Longitudinal Content Analysis, 1997-2003. In, Li X. *Internet Newspapers : Making of a Mainstream Medium*, London: Lawrence Erlbaum Associates, p.13-32.

Grossi, P., Junreuther, H. (2005) *Catastrophe Modelling: A New Approach to Managing Risk*, New York: Springer.

Guinau, M., Pallàs, R. and Vilaplana, J. M. (2005) 'A Feasible Methodology for Landslide Susceptibility Assessment in Developing Countries: A Case-Study of NW Nicaragua after Hurricane Mitch' *Engineering Geology,* 80**(3–4)**, p.316-327.

Guinau, M., Vilajosana, I. and Vilaplana, J. (2007) 'GIS-Based Debris Flow Source and Runout Susceptibility Assessment from Dem Data? A Case Study in NW Nicaragua' *Natural Hazards and Earth System Science,* 7**(6)**, p.703-716.

Guns, M. and Vanacker, V. (2014) 'Shifts in Landslide Frequency–Area Distribution after Forest Conversion in the Tropical Andes' *Anthropocene,* 6**(0)**, p.75-85.

Gutenberg, B. and Richter, C. F. (1944) 'Frequency of Earthquakes in California' *Bulletin of the Seismological Society of America,* 34**(4)**, p.185-188.

Guthrie, R. H. and Evans, S. G. (2004) 'Magnitude and Frequency of Landslides Triggered by a Storm Event, Loughborough Inlet, British Columbia' *Nat. Hazards Earth Syst. Sci.,* 4**(3)**, p.475-483.

Guthrie, R. H. and Evans, S. G. (2007) 'Work, Persistence, and Formative Events: The Geomorphic Impact of Landslides' *Geomorphology,* 88**(3–4)**, p.266-275.

Guthrie, R., Deadman, P., Cabrera, A. and Evans, S. (2008) 'Exploring the Magnitude–Frequency Distribution: A Cellular Automata Model for Landslides' *Landslides,* 5**(1)**, p.151-159.

Guzzetti, F. (2000) 'Landslide Fatalities and the Evaluation of Landslide Risk in Italy' *Engineering Geology,* 58**(2)**, p.89-107.

Guzzetti, F. (2006) *Landslide Hazard and Risk Assessment*, Ph.D. Thesis, Universität Bonn

Guzzetti, F. and Tonelli, G. (2004) 'Information System on Hydrological and Geomorphological Catastrophes in Italy (SICI): A Tool for Managing Landslide and Flood Hazards' *Nat. Hazards Earth Syst. Sci.,* 4**(2)**, p.213-232.

Guzzetti, F., Ardizzone, F., Cardinali, M., Galli, M., Reichenbach, P.and Rossi, M. (2008) 'Distribution of Landslides in the Upper Tiber River Basin, Central Italy' *Geomorphology,* 96**(1-2)**, p.105-122.

Guzzetti, F., Ardizzone, F., Cardinali, M., Rossi, M. and Valigi, D. (2009) 'Landslide Volumes and Landslide Mobilization Rates in Umbria, Central Italy' *Earth and Planetary Science Letters,* 279**(3)**, p.222-229.

Guzzetti, F., Cardinali, M. (1989) 'Carta Inventario dei Movimenti Franosi della Regione dell'Umbria ed aree Limitrofe' N.204, 2 sheets, Scale: 1:100 000, Umbria: CNR.

Guzzetti, F., Cardinali, M. and Reichenbach, P . (1994) 'The AVI Project: A Bibliographical and Archive Inventory of Landslides and Floods in Italy' *Environmental Management,* 18**(4)**, p.623-633.

Guzzetti, F., Cardinali, M., Reichenbach, P., Cipolla, F., Sebastiani, C., Galli, M. and Salvati, P . (2004a) 'Landslides Triggered by the 23 November 2000 Rainfall Event in the Imperia Province, Western Liguria, Italy' *Engineering Geology,* 73**(3-4)**, p.229-245.

Guzzetti, F., Carrara, A., Cardinali, M. and Reichenbach, P . (1999) 'Landslide Hazard Evaluation: A Review of Current Techniques and Their Application in a Multi-Scale Study, Central Italy' *Geomorphology,* 31**(1–4)**, p.181-216.

Guzzetti, F., Galli, M., Reichenbach, P., Ardizzone, F. and Cardinali, M. (2006a) 'Landslide Hazard Assessment in the Collazzone Area, Umbria, Central Italy' *Natural Hazards and Earth System Science,* 6**(1)**, p.115-131.

Guzzetti, F., Malamud, B. D., Turcotte, D. L. and Reichenbach, P . (2002) 'Power-Law Correlations of Landslide Areas in Central Italy' *Earth and Planetary Science Letters,* 195**(3–4)**, p.169-183.

Guzzetti, F., Mondini, A. C., Cardinali, M., Fiorucci, F., Santangelo, M. and Chang, K.-T. (2012) 'Landslide Inventory Maps: New Tools for an Old Problem' *Earth-Science Reviews,* 112**(1–2)**, p.42-66.

Guzzetti, F., Peruccacci, S., Rossi, M. and Stark, C. (2008b) 'The Rainfall Intensity–Duration Control of Shallow Landslides and Debris Flows: An Update' *Landslides,* 5**(1)**, p.3-17.

Guzzetti, F., Reichenbach, P.and Ghigi, S. (2004b) 'Rockfall Hazard and Risk Assessment Along a Transportation Corridor in the Nera Valley, Central Italy' *Environmental Management,* 34**(2)**, p.191-208.

Guzzetti, F., Reichenbach, P., Ardizzone, F., Cardinali, M. and Galli, M. (2006b) 'Estimating the Quality of Landslide Susceptibility Models' *Geomorphology,* 81**(1)**, p.166-184.

Guzzetti, F., Reichenbach, P., Cardinali, M., Ardizzone, F. and Galli, M. (2003) 'The Impact of Landslides in the Umbria Region, Central Italy' *Natural Hazards and Earth System Science,* 3**(5)**, p.469-486.

Guzzetti, F., Reichenbach, P., Cardinali, M., Galli, M. and Ardizzone, F. (2005) 'Probabilistic Landslide Hazard Assessment at the Basin Scale' *Geomorphology,* 72**(1-4)**, p.272-299.].

Haggett, P.and Chorley, R. J. (1969) *Network Analysis in Geography*, London: Edward Arnold.

Haghani, A. and Oh, S.-C. (1996) 'Formulation and Solution of a Multi-Commodity, Multi-Modal Network Flow Model for Disaster Relief Operations' *Transportation Research Part A: Policy and Practice,* 30**(3)**, p.231-250.

Haklay, M. and Weber, P . (2008) 'Openstreetmap: User-Generated Street Maps' *Pervasive Computing, IEEE,* 7**(4)**, p.12-18.

Han, Y. and Davidson, R. A. (2012) 'Probabilistic Seismic Hazard Analysis for Spatially Distributed Infrastructure' *Earthquake Engineering & Structural Dynamics,* 41**(15)**, p.2141-2158.

Hansen, R. and Bausch, D. (2007) 'A GIS-Based Methodology for Exporting the Hazards Us (HAZUS) Earthquake Model for Global Applications' [online] Available at: http://www.usehazus.com/docs/gis_global_hazus_paper.pdf [Accessed 27/04/2015].

Harp E. L., Jibson R. L. (1995) *Inventory of landslides triggered by the 1994 Northridge, California earthquake.* Open File Report 95–213, US Geological Survey.

Harp E. L., Jibson, R. L. (1996) 'Landslides triggered by the 1994 Northridge, California earthquake' *Seismological Society of America Bulletin,* **86**: S319– S332.

Harp, E. L., Jibson, R. W. and Dart, R. L (2013) The Effect of Complex Fault Rupture on the Distribution of Landslides Triggered by the 12 January 2010, Haiti Earthquake. In: Margottini, C., Canuti, P., Sassa, K. (eds.) *Landslide Science and Practice Volume 5: Complex Environment*, Berlin: Springer, p.157-161.

Harp, E. L., Keefer, D. K., Sato, H. P.and Yagi, H. (2011) 'Landslide Inventories: The Essential Part of Seismic Landslide Hazard Analyses' *Engineering Geology,* 122**(1–2)**, p.9-21.

Harp, E.L. and Keefer, D.K. (1990) 'Landslides triggered by the earthquake' in Rymer, M.J. and Ellsworth, W.L. (eds.) *The Coalinga, California, Earthquake of May 2, 1983*, U.S. Geological Survey Professional Paper 1487, 335–347, 1 pl., Scale 1:48,000.

Harp, E.L., Wilson, R.C., and Wieczorek, G.F. (1981) 'Landslides from the February 4, 1976, Guatemala earthquake' U.S. Geological Survey Professional Paper 1204-A, 35 p., 2 pl., Scale 1:50,000.

Harrell, F. E. (2001) *Regression Modeling Strategies : With Applications to Linear Models, Logistic Regression, and Survival Analysis*, London: Springer.

Hearn, G. J . (ed.) (2011) *Slope Engineering for Mountain Roads*, publication no.24, London: Geological Society.

Hearn, G., Hunt, T., Aubert, J., Howell, J. (2007) 'Landslide Impacts on the Road Network of Lao PDR and the Feasibility of Implementing a Slope Management Program' *Department for Industry and Development (DFID) Research For Development (R4D)* [online] Available at: http://www.dfid.gov.uk/R4D/PDF/Outputs/SeaCap/Seacp21-02.pdf [Accessed 29/02/2012].

Heath, T. L. (1931) *A Manual of Greek Mathematics*. Courier Corporation.

Heim, A. (1989 [1932].) *Landslides & Human Lives [Translated by Skermer, N. A.]*. Vancouver, B.C.: BiTech Publishers.

Hergarten, S. (2003) 'Landslides, Sandpiles, and Self-Organized Criticality' *Natural Hazards and Earth System Science,* 3**(6)**, p.505-514.

Hervás, J. (2013) 'Definition: Landslide' In Bobrowsky, P.T. (ed.) *Encyclopedia of Natural Hazards*, Netherlands: Springer.

Hintze, J. L. and Nelson, R. D. (1998) 'Violin Plots: A Box Plot-Density Trace Synergism' *The American Statistician,* 52**(2)**, p.181-184.

Hochrainer, S. (2006) *Macroeconomic Risk Management against Natural Disasters*. Springer.

Hofmeister, R. J. (2001) 'Slope Failures in Oregon GIS Inventory for Three 1996/97 Storm Events' [online] Available at: http://www.oregon.gov/dogami/pages/ landslide/ inventory/project.aspx [Accessed 27/04/2015].

Hong, Y., Adler, R., and Huffman, G. (2007) 'Use of satellite remote sensing data in the mapping of global landslide susceptibility' *Natural Hazards*, 43**(2)**, p.245-256.

Hopkins, D. J. and King, G. (2010) 'A Method of Automated Nonparametric Content Analysis for Social Science' *American Journal of Political Science,* 54**(1)**, p.229-247.

Hovius, N., Stark, C. P.and Allen, P.A. (1997) 'Sediment Flux from a Mountain Belt Derived by Landslide Mapping' *Geology,* 25**(3)**, p.231-234.

Hovius, N., Stark, C. P., Hao-Tsu, C. and Jiun-Chuan, L. (2000) 'Supply and Removal of Sediment in a Landslide-Dominated Mountain Belt: Central Range, Taiwan' *The Journal of Geology,* 108**(1)**, p.73-89.

Huang, R. Q. and Li, W. L. (2009) 'Analysis of the Geo-Hazards Triggered by the 12 May 2008 Wenchuan Earthquake, China' *Bulletin of Engineering Geology and the Environment,* 68**(3)**, p.363-371.

Huang, Z., Rosowsky, D. V. and Sparks, P.R. (2001) 'Long-Term Hurricane Risk Assessment and Expected Damage to Residential Structures' *Reliability Engineering & System Safety,* 74**(3)**, p.239-249.

Hungr, O., Corominas, J. and Eberhardt, E. (2005) 'Estimating Landslide Motion Mechanism, Travel Distance and Velocity' *Landslide Risk Management*, p.99-128.

Hungr, O., Evans, S. and Hazzard, J. (1999) 'Magnitude and Frequency of Rock Falls and Rock Slides Along the Main Transportation Corridors of Southwestern British Columbia' *Canadian Geotechnical Journal,* 36**(2)**, p.224-238.

Hungr, O., Leroueil, S. and Picarelli, L. (2014) 'The Varnes Classification of Landslide Types, an Update' *Landslides,* 11**(2)**, p.167-194.

Hungr, O., McDougall, S., Wise, M. and Cullen, M. (2008) 'Magnitude–Frequency Relationships of Debris Flows and Debris Avalanches in Relation to Slope Relief' *Geomorphology,* 96**(3–4)**, p.355-365.

Hurford, A.P., Priest, S.J., Parker, D. J. and Lumbroso, D. M. (2012) 'The effectiveness of extreme rainfall alerts in predicting surface water flooding in England and Wales' *International Journal of Climatology*, 32**(11)**: 1768–1774.

Hurst, M. D., Ellis, M. A., Royse, K. R., Lee, K. A. and Freeborough, K. (2013) 'Controls on the Magnitude-Frequency Scaling of an Inventory of Secular Landslides' *Earth Surf. Dynam.,* 1**(1)**, p.67-78.

Ibsen, M.-L. and Brunsden, D. (1996) 'The Nature, Use and Problems of Historical Archives for the Temporal Occurrence of Landslides, with Specific Reference to the South Coast of Britain, Ventnor, Isle of Wight' *Geomorphology,* 15**(3–4)**, p.241-258.

IECA (International Erosion Control Association) (2012) 'IECA Photo Gallery: Asia Trip Plants Seeds for Chapter Growth' [online] Available at: http://www.ieca.org/photogallery /AsiaTripPhotos.asp?printerfriendly=true& [Accessed 06/03/2012].

IRPI WebMapping (2015) 'Map of Archive and Geomorophological Inventories for the Perugia Region' [online] Available at: http://webmap.irpi.cnr.it/ [Accessed 27/04/2015].

ISPRA (Instituto Superiore per la Protezione e la Ricerca Ambientale (2011) 'Inventory of Landslide Phenomena in Italy Project' [online] Available at: http://www.isprambiente.gov.it/site/en-GB/Projects/IFFI_Project/default.html [Accessed 02/05/2012].

Jaboyedoff, M., Oppikofer, T., Abellán, A., Derron, M.-H., Loye, A., Metzger, R. and Pedrazzini, A. (2012) 'Use of LiDAR in Landslide Investigations: A Review' *Natural Hazards,* 61**(1)**, p.5-28.

Jaiswal, P., van Westen, C. and Jetten, V. (2011) 'Quantitative Assessment of Landslide Hazard Along Transportation Lines Using Historical Records' *Landslides,* 8**(3)**, p.279-291.

Jaiswal, P., van Westen, C. J. and Jetten, V. (2010b) 'Quantitative Landslide Hazard Assessment Along a Transportation Corridor in Southern India' *Engineering Geology,* 116**(3–4)**, p.236-250.

Jaiswal, P., Westen, C. J. v. and Jetten, V. (2010a) 'Quantitative Assessment of Direct and Indirect Landslide Risk Along Transportation Lines in Southern India' *Natural Hazards and Earth System Sciences*, 10, p.1253-1267.

Jakob, M., Hungr, O. and Jakob, D. M. (2005) *Debris-Flow Hazards and Related Phenomena*. Berlin: Springer.

James, F. (2006) *Statistical Methods in Experimental Physics*. Singapore: World Scientific.

James, G. A., Cliff, A. D., Haggett, P.and Ord, J. K. (1970) 'Some Discrete Distributions for Graphs with Applications to Regional Transport Networks' *Geografiska Annaler. Series B, Human Geography,* 52**(1)**, p.14-21.

Jarman, D. (2006) 'Large Rock Slope Failures in the Highlands of Scotland: Characterisation, Causes and Spatial Distribution' *Engineering Geology,* 83**(1)**, p.161-182.

Jenelius, E. (2009) 'Network Structure and Travel Patterns: Explaining the Geographical Disparities of Road Network Vulnerability' *Journal of Transport Geography,* 17**(3)**, p.234-244.

Jenelius, E. and Mattsson, L.-G. (2012) 'Road Network Vulnerability Analysis of Area-Covering Disruptions: A Grid-Based Approach with Case Study' *Transportation Research Part A: Policy and Practice,* 46**(5)**, p.746-760.

Jenelius, E., Petersen, T. and Mattsson, L.-G. (2006) 'Importance and Exposure in Road Network Vulnerability Analysis' *Transportation Research Part A: Policy and Practice,* 40**(7)**, p.537-560.

Jensen, J. (2000) *Remote Sensing of the Environment: An Earth Resource Perspective*, New Jersey: Prentice Hall.

Jibson, R. W. (1992) 'Chapter 5: The Mameyes, Puerto Rico, Landslide Disaster of October 7, 1985' *Reviews in Engineering Geology,* 9, p.37-54.

Jibson, R. W., Harp, E., L., and Michael, J. A. (1998) 'A Method for Producing Digital Probabilistic Seismic Landslide Hazard Maps: An Example from the Los Angeles, California, Area' Open file report 98-113, United States Geological Survey.

Jones D.K.C. and Lee E.M. (1994) *Landsliding in Great Britain,* London:Department of the Environment, London, 390 p.

Jones, E., Oliphant, T. and Peterson, P . (2014) '{Scipy}: Open Source Scientific Tools for {Python}' [online] Available at: http://www.scipy.org/ [Accessed 27/04/2015].

Jordan, L., Stallins, A., Stokes IV, S., Johnson, E. and Gragg, R. (2011) 'Citizen Mapping and Environmental Justice: Internet Applications for Research and Advocacy' *Environmental Justice,* 4**(3)**, p.155-162.

Kagan, Y. Y. and Knopoff, L. (1980) 'Spatial Distribution of Earthquakes: The Two-Point Correlation Function' *Geophysical Journal International,* 62**(2)**, p.303-320.

Kang, Y.-J. and Wen, Y. (2000) 'Minimum Life-Cycle Cost Structural Design against Natural Hazards' University of Illinois Engineering Experiment Station. College of Engineering. University of Illinois at Urbana-Champaign.

Kansky, K. (1963) *Structure of Transportation Networks*, Chicago: University of Chicago Press.

Kasperson, R. E., Renn, O., Slovic, P., Brown, H. S., Emel, J., Goble, R., Kasperson, J. X. and Ratick, S. (1988) 'The Social Amplification of Risk - a Conceptual-Framework' *Risk Analysis,* 8**(2)**, p.177-187.

Katz, O. and Aharonov, E. (2006) 'Landslides in Vibrating Sand Box: What Controls Types of Slope Failure and Frequency Magnitude Relations?' *Earth and Planetary Science Letters,* 247**(3–4)**, p.280-294.

Katz, O., Morgan, J. K., Aharonov, E. and Dugan, B. (2014) 'Controls on the Size and Geometry of Landslides: Insights from Discrete Element Numerical Simulations' *Geomorphology,* 220**(0)**, p.104-113.

Keaton, J. and DeGraff, J. (1996) 'Surface Observation and Geologic Mapping' In: Turner, K. and Schuster, R. (eds.) *Landslides Investigation and Mitigation Transportation Research Board Special Report 247*, Washington DC: Transportation Research Board, p.178-230

Keefer, D. K. (2000) 'Statistical Analysis of an Earthquake-Induced Landslide Distribution -- the 1989 Loma Prieta, California Event' *Engineering Geology,* 58**(3-4)**, p.231-249.

Keefer, D. K. (2002) 'Investigating landslides caused by earthquakes–a historical review' *Surveys in Geophysics*, 23**(6)**, p.473-510.

Khater, M., Scawthorn, C. and Johnson, J. 2003. Loss Estimation. *In,* eds.) (Scawthorn and Wai-Fah)*Earthquake Engineering Handbook*. Boca Raton: CRC Press. p.1-18

Kircher, C. A., Whitman, R. V. and Holmes, W. T. (2006) 'Hazus Earthquake Loss Estimation Methods' *Natural Hazards Review,* 7**(2)**, p.45-59.

Kirilenko, A. and Stepchenkova, S. (2012) 'Climate Change Discourse in Mass Media: Application of Computer-Assisted Content Analysis' *Journal of Environmental Studies and Sciences,* 2**(2)**, p.178-191.

Kirschbaum, D., Adler, R., Hong, Y., Hill, S. and Lerner-Lam, A. (2010) 'A Global Landslide Catalog for Hazard Applications: Method, Results, and Limitations' *Natural Hazards,* 52**(3)**, p.561-575.

Klar, A., Aharonov, E., Kalderon-Asael, B., and Katz, O. (2011) 'Analytical and observational relations between landslide volume and surface area' *Journal of Geophysical Research: Earth Surface, 116***(F2)**

Klose, M., Damm, B. and Terhorst, B. (2014) 'Landslide Cost Modelling for Transportation Infrastructures: A Methodological Approach' *Landslides*, p.1-14.

Korup, O., Clague, J. J., Hermanns, R. L., Hewitt, K., Strom, A. L. and Weidinger, J. T. (2007) 'Giant Landslides, Topography, and Erosion' *Earth and Planetary Science Letters,* 261**(3–4)**, p.578-589.

Kuwata, Y. and Takada, S. (2004) 'Effective Emergency Transportation for Saving Human Lives' *Natural Hazards,* 33**(1)**, p.23-46.

LAMPRE (2014) 'Landslide Inventory Map' [online] Available at: http://www.lampre-project.eu/index.php?option=com_k2&view=item&layout=item&id=169&Itemid=739. [Accessed 27/04/2015].

Landsea, C. W., Anderson, C., Charles, N., Clark, G., Dunion, J., Fernandez-Partagas, J., Hungerford, P., Neumann, C. and Zimmer, M. (2004) 'The Atlantic Hurricane Database Re-Analysis Project: Documentation for the 1851–1910 Alterations and Additions to the Hurdat Database' *Hurricanes and Typhoons: Past, Present and Future*, p.177-221.

Lanteri, L., and Colombo, A. (2013) The integration between satellite data and conventional monitoring system in order to update the Arpa Piemonte landslide inventory. In . In: Margottini, C., Canuti, P., Sassa, K. (eds.) *Landslide science and practice,* p.135-140. Springer Berlin Heidelberg.

Larsen, M. C. and Parks, J. E. (1997) 'How Wide Is a Road? The Association of Roads and Mass-Wasting in a Forested Montane Environment' *Earth Surface Processes and Landforms,* 22**(9)**, p.835-848.

Lee, C., Wang, I. (2011) 'Analysis of Highway Slope Failure by an Application of the Stereographic Projection' *World Academy of Science, Engineering and Technology*, 75, p.515-523.

Legros, F. (2002) 'The Mobility of Long-Runout Landslides' *Engineering Geology,* 63**(3–4)**, p.301-331.

Lehmann, P., and Or, D. (2012) 'Hydromechanical triggering of landslides: From progressive local failures to mass release' *Water Resources Research*, *48***(3)**

LexisNexis Academic (2014) [online] Available at: http://academic.lexisnexis.com [Accessed 2/9/14].

Li, X., Lu, L., Cheng, G. and Xiao, H. (2001) 'Quantifying Landscape Structure of the Heihe River Basin, North-West China Using Fragstats' *Journal of Arid Environments,* 48**(4)**, p.521-535.

Liew, C. (2001) 'Microwave Remote Sensing' *Principles of Remote Sensing Tutorial from Centre for Remote Imaging Sensing and Processing* [online] Available at: http://www.crisp.nus.edu.sg/~research/tutorial/mw.htm [Accessed 02/03/2012].

Lilliefors, H. W. (1967) 'On the Kolmogorov-Smirnov Test for Normality with Mean and Variance Unknown' *Journal of the American Statistical Association,* 62**(318)**, p.399-402.

Limpert, E., Stahel, W. A. and Abbt, M. (2001) 'Log-Normal Distributions across the Sciences: Keys and Clues' *BioScience,* 51**(5)**, p.341-352.

Llasat, M. C., Llasat-Botija, M. and López, L. (2009) 'A Press Database on Natural Risks and Its Application in the Study of Floods in Northeastern Spain' *Natural Hazards and Earth System Sciences,* 9**(6)**, p.2049-2061.

Lo, C.-M., Lee, C.-F., Chou, H.-T. and Lin, M.-L. (2014) 'Landslide at Su-Hua Highway 115.9k Triggered by Typhoon Megi in Taiwan' *Landslides,* 11**(2)**, p.293-304.

Lombardo, U. (2014) 'Quantitative Morphometric Analysis of Lakes Using GIS: Rectangularity R, Ellipticity E, Orientation O, and the Rectangularity Vs. Ellipticity Index, Rei' *Cartography and Geographic Information Science,* 41**(4)**, p.340-347.

Longley, P., Goodchild, M., Maguire, D., Rhind, D. (2010) *Geographyic Information Systems and Science (3rd Edition)*, Chichester: John Wiley and Sons.

Lu, P., Bai, S., and Casagli, N. (2014) 'Investigating spatial patterns of persistent scatterer interferometry point targets and landslide occurrences in the Arno River Basin' *Remote Sensing*, 6**(8)**, p.6817-6843.

Lu, P., Casagli, N., Catani, F. and Tofani, V. (2012) 'Persistent Scatterers Interferometry Hotspot and Cluster Analysis (Psi-Hca) for Detection of Extremely Slow-Moving Landslides' *International Journal of Remote Sensing,* 33**(2)**, p.466-489.

Maclachlan, J. C. and Eyles, C. H. (2013) 'Quantitative Geomorphological Analysis of Drumlins in the Peterborough Drumlin Field, Ontario, Canada' *Geografiska Annaler: Series A, Physical Geography,* 95**(2)**, p.125-144.

Malamud, B. D. and Turcotte, D. L. (1999) 'Self-Organized Criticality Applied to Natural Hazards' *Natural Hazards,* 20**(2)**, p.93-116.

Malamud, B. D., Mihir, M., Reichenbach, P., Rossi, M. (2014) 'LAMPRE Deliverable 6.3 Report on Standards for Landslide Sucsceptibility Modelling and Terrain Zonations' [online] Available at: http://www.lampre-project.eu/index.php?option=com_phocadownload&view=category&id=27:wp6-preparedeness-prevention-recovery-reconstruction&Itemid=525 [Accessed 27/04/2015].

Malamud, B. D., Turcotte, D. L., Guzzetti, F. and Reichenbach, P . (2004a) 'Landslide Inventories and Their Statistical Properties' *Earth Surface Processes and Landforms,* 29**(6)**, p.687-711.

Malamud, B. D., Turcotte, D. L., Guzzetti, F., Reichenbach, P . (2004b) 'Landslides, earthquakes and erosion' *Earth and Planetary Science Letters*, 229**(1-2)**, p.45-59.

Malamud, B.D. (2004) Tails of natural hazards: *Physics World*, 17**(8),** p.31**–**35.

Marchesini, I., Cencetti, C. and De Rosa, P . (2009) 'A Preliminary Method for the Evaluation of the Landslides Volume at a Regional Scale' *GeoInformatica,* 13**(3)**, p.277-289.

Martel, S. J. (2004) 'Mechanics of landslide initiation as a shear fracture phenomenon' *Marine Geology*, 203**(3)**, p.319-339.

Martha, T. R., Kerle, N., Jetten, V., van Westen, C. J., and Kumar, K. V. (2010) Characterising spectral, spatial and morphometric properties of landslides for semi-automatic detection using object-oriented methods. *Geomorphology*,*116*(1), 24-36.

Martha, T., Roy, P., Govindharaj, K. B., Kumar, K. V., Diwakar, P.G. and Dadhwal, V. K. (2015) 'Landslides Triggered by the June 2013 Extreme Rainfall Event in Parts of Uttarakhand State, India' *Landslides,* 12**(1)**, p.135-146.

McGuire, R. K. (1995) 'Probabilistic Seismic Hazard Analysis and Design Earthquakes: Closing the Loop' *Bulletin of the Seismological Society of America,* 85**(5)**, p.1275-1284.

McKean, J. and Roering, J. (2004) 'Objective Landslide Detection and Surface Morphology Mapping Using High-Resolution Airborne Laser Altimetry' *Geomorphology,* 57**(3–4)**, p.331-351.

McLellan, T. and Endler, J. A. (1998) 'The Relative Success of Some Methods for Measuring and Describing the Shape of Complex Objects' *Systematic Biology,* 47**(2)**, p.264-281.

Meltwater (2014) Meltwater News [online] Available at: http://www.meltwater.com/products/meltwater-news/ [Accessed 02 Sep 2014].

Merz, B., Kreibich, H., Thieken, A. and Schmidtke, R. (2004) 'Estimation Uncertainty of Direct Monetary Flood Damage to Buildings' *Natural Hazards and Earth System Science,* 4**(1)**, p.153-163.

Met Office (2011) 'Met Office Weather Observations Website (WOW' [online] Available at: http://wow.metoffice.gov.uk/ [Accessed 27/04/2015].

Met Office. (2013) 'Natural Hazards Partnership: Hazard Impact Model' [online] Available at: http://www.metoffice.gov.uk/nhp/hazard-impact-model [Accessed 27/04/2015].

Metternicht, G., Hurni, L. and Gogu, R. (2005) 'Remote Sensing of Landslides: An Analysis of the Potential Contribution to Geo-Spatial Systems for Hazard Assessment in Mountainous Environments' *Remote Sensing of Environment,* 98**(2–3)**, p.284-303.

Meunier, P., Hovius, N. and Haines, A. J. (2007) 'Regional Patterns of Earthquake-Triggered Landslides and Their Relation to Ground Motion' *Geophysical Research Letters,* 34**(20)**

Meunier, P., Hovius, N., and Haines, J. A. (2008) 'Topographic site effects and the location of earthquake induced landslides' *Earth and Planetary Science Letters*, 275**(3)**, p.221-232.

Meunier, P., Uchida, T. and Hovius, N. (2013) 'Landslide Patterns Reveal the Sources of Large Earthquakes' *Earth and Planetary Science Letters,* 363**(0)**, p.27-33.

Meyer, N. K., Schwanghart, W., Korup, O., Nadim, F. (2015) 'Roads and Risk – Traffic Detours from Debris Flows in Southern Norway' *Natural Hazards and Earth System Sciences (Discussions stage)*, DOI: 10.5194/nhessd-2-6623-2014.

Michon, G. (2015) 'Final Answers: Perimeter of an Ellipse' [online] Available at: http://www.numericana.com/answer/ellipse.htm [Accessed 27/04/2015].

Milledge, D. G., Bellugi, D., McKean, J. A., Densmore, A. L., and Dietrich, W. E. (2014) 'A multidimensional stability model for predicting shallow landslide size and shape across landscapes' *Journal of Geophysical Research: Earth Surface*, *119***(11)**, p.2481-2504.

Milledge, D. G., Griffiths, D. V., Lane, S. N. and Warburton, J. (2012) 'Limits on the Validity of Infinite Length Assumptions for Modelling Shallow Landslides' *Earth Surface Processes and Landforms,* 37**(11)**, p.1158-1166.

Miller, D. J., and Burnett, K. M. (2007) 'Effects of forest cover, topography, and sampling extent on the measured density of shallow, translational landslides' *Water Resources Research*, *43***(3)**.

Miller, R. S., Franchi, M. and Flor, R. (2004) 'Grass GIS Manual: R. Drain' [online] Available at: http://grass.osgeo.org/grass64/manuals/r.drain.html [Accessed 27/04/2015].

Miller, S., Brewer, T. and Harris, N. (2009) 'Rainfall Thresholding and Susceptibility Assessment of Rainfall-Induced Landslides: Application to Landslide Management in St Thomas, Jamaica' *Bulletin of Engineering Geology and the Environment,* 68**(4)**, p.539-550.

Moeller, S. D. (2006) "Regarding the Pain of Others': Media, Bias and the Coverage of International, Disasters' *Journal of International Affairs,* 59**(2)**, p.173.

Moine, M., Puissant, A., and Malet, J. P . (2009) 'Detection of landslides from aerial and satellite images with a semi-automatic method. Application to the Barcelonnette basin (Alpes-de-Hautes-Provence, France)' In Malet, J-P., Remaitre, A., Bogaard, T. (eds.) *Landslide processes-from geomorphologic mapping to dynamic modelling,* p.63-68.

Mondini, A. C. and Chang, K.-t. (2014) 'Combining Spectral and Geoenvironmental Information for Probabilistic Event Landslide Mapping' *Geomorphology,* 213**(0)**, p.183-189.

Mondini, A. C., Chang, K.-T. and Yin, H.-Y. (2011a) 'Combining Multiple Change Detection Indices for Mapping Landslides Triggered by Typhoons' *Geomorphology,* 134**(3–4)**, p.440-451.

Mondini, A. C., Guzzetti, F., Reichenbach, P., Rossi, M., Cardinali, M. and Ardizzone, F. (2011b) 'Semi-Automatic Recognition and Mapping of Rainfall Induced Shallow Landslides Using Optical Satellite Images' *Remote Sensing of Environment,* 115**(7)**, p.1743-1757.

Mondini, A., Viero, A., Cavalli, M., Marchi, L., Herrera, G. and Guzzetti, F. (2014) 'Comparison of Event Landslide Inventories: The Pogliaschina Catchment Test Case, Italy' *Nat Hazards Earth Syst Sci Discuss,* 2**(**p.1093-1125.

Montgomery, D. R. and Dietrich, W. E. (1994) 'A Physically Based Model for the Topographic Control on Shallow Landsliding' *Water resources research,* 30**(4)**, p.1153-1171.

Moon, A.T., Wilson, R.A., Flentje, P.N. (2005), Developing and using landslide size frequency models' In: *Proceedings of the International Conference on Landslide Risk Management, 18th Annual Vancouver Geotechnical Society Symposium*, Vancouver, 31 May - 4 June 2005, p.589-598.

Mooney, P.and Corcoran, P . (2013) 'Has Openstreetmap a Role in Digital Earth Applications?' *International Journal of Digital Earth,* 7**(7)**, p.534-553.

Moores, A.J. and Rees, JG. (eds.) (2011) *UK Flood and Coastal Erosion Risk Management Research Strategy*. Living with Environmental Change. [online] Available at: http://www.lwec.org.uk/our-work/uk-first-flood-research-strategy [Accessed 02/09/2014].

Mota, B. (2014) 'Landslide inventory map for Messina (Italy) using coarse-resolution remote sensing imagery' in Aridzzone, F., Del Ventisette, C., Drake, N., Fiorucci, F., Malamud, B., Mondini, A., Mota, B. (eds.) *LAMPRE Report on Methods for the Produciton of Event*

Landslide Inventory Maps Exploiting EO Imagery [online] Available at: http://www.lampre-project.eu/index.php?option=com_phocadownload&view=category&id=26:wp5-triggered-event-landslides&Itemid=525 [Accessed 27/04/2015].

Muenchow, J., Brenning, A., and Richter, M. (2012) 'Geomorphic process rates of landslides along a humidity gradient in the tropical Andes' *Geomorphology*,139, p.271-284.

Murillo-García, F. G., Alcántara-Ayala, I., Ardizzone, F., Cardinali, M., Fiourucci, F., & Guzzetti, F. (2014) 'Satellite stereoscopic pair images of very high resolution: a step forward for the development of landslide inventories' *Landslides*, 12**(2)**, p.227-291.

Murty, C. V. R., Raghukanth, S. T. G., Menon, A., Goswami, R., Vijayanarayanan, A. R., Gandhi, S. R., Satyanarayana, K. N., Sheth, A., Jaiswal, A., Kaushik, H. B., Dasgupta, K., Chaurasia, A., Roorkee, S. B., Roy, D., Pradeep, R. P., Mitra, K., Rai, D. C., Mondal, G., Singhal, S., Parool, P.and Pradhan, P . (2012) 'The M_w 6.9 Sikkim-Nepal Border Earthquake of September 18, 2011' In: *EERI Special Earthquake Report*. [online] Available at: https://www.eeri.org/wp-content/uploads/Sikkim-EQ-report-02-07.pdf [Accessed 27/04/2015].

Musson, R. (2000) 'The Use of Monte Carlo Simulations for Seismic Hazard Assessment in the UK' *Annals of Geophysics*, 43**(1**), p.1-9.

Myung, I. J. (2003) 'Tutorial on Maximum Likelihood Estimation' *Journal of Mathematical Psychology,* 47**(1)**, p.90-100.

NCSE. (2007) 'Nineteenth Century Serials Edition' [online] Available at: http://www.ncse.ac.uk/index.html [Accessed 27/04/2015].

Neighbors, C., Cochran, E., Caras, Y. and Noriega, G. (2012) 'Sensitivity Analysis of FEMA HAZUS Earthquake Model: Case Study from King County, Washington' *Natural Hazards Review,* 14**(2)**, p.134-146.

Newman, M. E. J. (2005) 'Power Laws, Pareto Distributions and Zipf's Law' *Contemporary Physics*, 46**(5)**, 323-351.

Nichol, J. and Wong, M. S. (2005) 'Satellite Remote Sensing for Detailed Landslide Inventories Using Change Detection and Image Fusion' *International Journal of Remote Sensing,* 26**(9)**, p.1913-1926.

Nichol, J. and Wong, M. S. (2007) 'Remote Sensing of Urban Vegetation Life Form by Spectral Mixture Analysis of High-Resolution Ikonos Satellite Images' *International Journal of Remote Sensing,* 28**(5)**, p.985-1000.

Niethammer, U., James, M. R., Rothmund, S., Travelletti, J. and Joswig, M. (2012) 'Uav-Based Remote Sensing of the Super-Sauze Landslide: Evaluation and Results' *Engineering Geology,* 128**(0)**, p.2-11.

Nojima, N., Sugito, M. (2000) 'Simulation and Evaluation of Post-Earthquake Functional Performance of Transportation Network' *Proceedings of the 12th World Conference on Earthquake Engineering*, Auckland, New Zealand 30 January – 4 February 2000, New Zealand Society For Earthquake Engineering.

OCHA (United Nations Office for the Coordination of Humanitarian Affairs) (2011) 'Sikkim Earthquake, India. Situation Report Number 3, 22 September 2011' [online] Available

at: http://reliefweb.int/sites/reliefweb.int/files/resources/Sikkim%20India%20EQ%20 sitrep,%20no.%203,%2023%20Sept%202011.pdf [Accessed 25/10/2011].

OED Online (2014) 'Oxford English Dictionary' *Oxford University Press* [online] Available at: http://www.oed.com/ [Accessed 2/9/14.

Oh, H. J., and Pradhan, B. (2011) 'Application of a neuro-fuzzy model to landslide-susceptibility mapping for shallow landslides in a tropical hilly area' *Computers & Geosciences*, 37**(9)**, p.1264-1276.

Ortigão, J.A. R. (2004) 'Selection of Stabilisation Design' *In* Ortigão, J.A. R., Sayao, A. S. F. J. (eds.) *Handbook of Slope Stabilisation*, Berlin: Springer-Verlag, p.137- 147.

OSM (2014) 'Open Street Map Map Portal' [online] Available at: http://www.openstreetmap.org/#map=5/51.500/-0.100 [Accessed 27/04/2015].

Oven, K. J. (2009) *Landscape, Livelihoods and Risk: Community Vulnerability to Landslides in Nepal*, Ph.D. Thesis, University of Durham.

Oxford Dictionary (2015a) 'Database' [online] Available at: http://www.oxforddictionaries.com/definition/english/database [Accessed 27/04/2015].

Oxford Dictionary (2015b) 'Inventory' [online] Available at: http://www.oxforddictionaries.com/definition/english/inventory [Accessed 27/04/2015].

Parise, M. and Guzzi, R. (1992) 'Volume and Shape of the Active and Inactive Parts of the Slumgullion Landslide, Hinsdale County, Colorado' US Geological Survey.

Parise, M. and Jibson, R. W. (2000) 'A Seismic Landslide Susceptibility Rating of Geologic Units Based on Analysis of Characteristics of Landslides Triggered by the 17 January, 1994 Northridge, California Earthquake' *Engineering Geology,* 58**(3–4)**, p.251-270.

Parker, R. N., Densmore, A. L., Rosser, N. J., de Michele, M., Li, Y., Huang, R., Whadcoat, S. and Petley, D. N. (2011) 'Mass Wasting Triggered by the 2008 Wenchuan Earthquake Is Greater Than Orogenic Growth' *Nature Geosciences,* 4**(7)**, p.449-452.

Parry, S., Marsh, T. and Kendon, M. (2013) '2012: From Drought to Floods in England and Wales' *Weather,* 68**(10)**, p.268-274.

Parzen, E. (1962) 'On Estimation of a Probability Density Function and Mode' *The Annals of Mathematical Statistics,* 33**(3)**, p.1065-1076.

Peduzzi, P.and Herold, H. D. (2005) 'Mapping Disastrous Natural Hazards Using Global Datasets' *Natural Hazards,* 35**(2)**, p.265-289.

Pelletier, J. D., Malamud, B. D., Blodgett, T. and Turcotte, D. L. (1997) 'Scale-Invariance of Soil Moisture Variability and Its Implications for the Frequency-Size Distribution of Landslides' *Engineering Geology,* 48**(3–4)**, p.255-268.

Pennington, C. V. L. and Harrison, A. (2013) '2012 - Landslide Year' *In: Geoscientist*, 10 - 15. London: The Geological Society of London.

Pennington, C. V. L., Freeborough, K., Dashwood, C., Dijkstra, T. and Lawrie, K. (2015) 'The National Landslide Database of Great Britain: Acquisition, Communication and the

Role of Social Media.' *Geomorphology*, Advance article in press, doi:10.1016/j.geomorph. 2015.03.013.

Pennington, C., Dijkstra, T., Lark, M., Dashwood, C., Harrison, A. and Freeborough, K. (2014) Antecedent Precipitation as a Potential Proxy for Landslide Incidence in South West United Kingdom. In: Sassa, K., Canuti P.and Yin Y. (eds.) *Landslide Science for a Safer Geoenvironment Part V*. Switzerland: Springer International Publishing. p.253-259

Pennington, C., Foster, C., Chambers, J. and Jenkins, G. (2009) 'Landslide Research at the British Geological Survey: Capture, Storage and Interpretation on a National and Site-Specific Scale' *Acta Geologica Sinica - English Edition,* 83**(5)**, p.991-999.

Peters, H. P., Brossard, D., De Cheveigné, S., Dunwoody, S., Kallfass, M., Miller, S. and Tsuchida, S. (2008) 'Interactions with the Mass Media' *Science,* 321**(5886)**, p.204.

Petley, D. (2011) 'Landslide Protection, New Zealand Style (photo)' [online] Available at: http://blogs.agu.org/landslideblog/2011/07/31/landslide-protection-new-zealand-style/ [Accessed 01/05/2012].

Petley, D. (2012) 'Global Patterns of Loss of Life from Landslides' *Geology,* 40**(10)**, p.927-930.

Petley, D., Dunning, S., Rosser, N. (2005) 'The Analysis of Global Landslide Risk Through the Creation of a Database of Worldwide Landslide Fatalities' in Hungr, O., Fell, R., Couture, R., Eberhardt (eds.) *Landslide Risk Management*, London: Taylor and Francis, p.367-374.

Petley, D., Hearn, G., Hart, A., Rosser, N., Dunning, S., Oven, K. and Mitchell, W. (2007) 'Trends in Landslide Occurrence in Nepal' *Natural Hazards,* 43**(1)**, p.23-44.

Petrova, E. (2011) 'Critical Infrastructure in Russia: Geographical Analysis of Accidents Triggered by Natural Hazards' *Environmental engineering and management journal,* 10**(1)**, p.53-58.

Petrucci, O., Pasqua, A. A. and Gullà, G. (2010) 'Landslide Damage Assessment Using the Support Analysis Framework (Saf): The 2009 Landsliding Event in Calabria (Italy)' *Advances in Geoscience,* 26, p.13-17.

Picarelli, L., Oboni, F., Evans, S.G., Mostyn,G., Fell, R. (2005) 'Hazard characterization and quantification' In Hungr, O., Fell, R. Couture, R., Eberthardt E. (eds.), *Landslide Risk Management*, London: Taylor and Francis, p.27–61

Piegari, E., Maio, R. D., and Milano, L. (2009) 'Characteristic scales in landslide modelling' *Nonlinear Processes in Geophysics*, *16***(4)**, p.515-523.

Pinelli, J., Simiu, E., Gurley, K., Subramanian, C., Zhang, L., Cope, A., Filliben, J. and Hamid, S. (2004) 'Hurricane Damage Prediction Model for Residential Structures' *Journal of Structural Engineering,* 130**(11)**, p.1685-1691.

Pourghasemi, H. R., Moradi, H. R., Aghda, S. F., Sezer, E. A., Jirandeh, A. G., and Pradhan, B. (2014) 'Assessment of fractal dimension and geometrical characteristics of the landslides identified in North of Tehran, Iran' *Environmental earth sciences*, 71**(8)**, p.3617-3626.

Pradhan, B. and Lee, S. (2010) 'Regional Landslide Susceptibility Analysis Using Back-Propagation Neural Network Model at Cameron Highland, Malaysia' *Landslides,* 7**(1)**, p.13-30.

Press, W. H. (2002) *Numerical Recipes in C++ : The Art of Scientific Computing*. Cambridge: Cambridge University Press.

Public Health England (2014) 'Public Health England' [online] Available at: https://www.gov.uk/government/organisations/public-health-england. [Accessed 02/09/2014].

Quarantelli, E. L. (1996) 'Local Mass Media Operations in Disasters in the USA' *Disaster Prevention and Management: An International Journal,* 5**(5)**, p.5-10.

Quinn, P.E., Hutchinson, D. J., Diederichs, M. S. and Rowe, R. K. (2010) 'Regional-Scale Landslide Susceptibility Mapping Using the Weights of Evidence Method: An Example Applied to Linear Infrastructure' *Canadian Geotechnical Journal,* 47**(8)**, p.905-927.

Quinn, P.E., Hutchinson, D. J., Diederichs, M. S. and Rowe, R. K. (2011) 'Characteristics of Large Landslides in Sensitive Clay in Relation to Susceptibility, Hazard, and Risk' *Canadian Geotechnical Journal,* 48**(8)**, p.1212-1232.

Raines, G. L. (2002) 'Description and Comparison of Geologic Maps with Fragstats—a Spatial Statistics Program' *Computers & Geosciences,* 28**(2)**, p.169-177.

Raška, P., Zábranský, V., Dubišar, J., Kadlec, A., Hrbáčová, A. and Strnad, T. (2014) 'Documentary Proxies and Interdisciplinary Research on Historic Geomorphologic Hazards: A Discussion of the Current State from a Central European Perspective' *Natural Hazards,* 70**(1)**, p.705-732.

Rau, J., Jhan, J., Lo, C. and Lin, Y. (2011) 'Landslide Mapping Using Imagery Acquired by a Fixed-Wing Uav' *In: Proceedings of the International Conference on Unmanned Aerial Vehicle in Geomatics (UAV-g)*, 6.

Ray, P.C., Parvaiz, I., Jayangondaperumal, R., Thakur, V., Dadhwal, V. and Bhat, F. (2009) 'Analysis of Seismicity-Induced Landslides Due to the 8 October 2005 Earthquake in Kashmir Himalaya' *Current science,* 97, p.1742-1751.

Razak, K. A., and Brinksma, H. (2014) *Airborne laser scanning for forested landslides investigation in temperate and tropical environments* (Vol. 244) Utrecht University/University of Twente.

Reggiani, A., Nijkamp, P.and Lanzi, D. 'Transport Resilience and Vulnerability: The Role of Connectivity' *Transportation Research Part A: Policy and Practice*, Advance Article in Press, doi:10.1016/j.tra.2014.12.012.

Reichenbach, P., Ardizzone, F., cardinali, M., Galli, M., Guzzetti, F., Salvati, P . (2003) 'Landslide Events and their impact on the transportation network in the Umbria region, Central Italy' *Proceedings of the 4th EGS Plinius Conference on Mediterranean Storms*, Mallorca, Spain, October 2002.

Remondo, J., Bonachea, J. and Cendrero, A. (2008) 'Quantitative Landslide Risk Assessment and Mapping on the Basis of Recent Occurrences' *Geomorphology,* 94**(3–4)**, p.496-507.

Remzi, E. K. E. R., and Aydin, A. (2014), 'Assessment of forest road conditions in terms of landslide susceptibility: a case study in Yığılca Forest Directorate (Turkey)' *Turkish Journal of Agriculture and Forestry*, 38**(2)**, p.281-290.

Reneau, S. L., and Dietrich, W. E. (1987) 'Size and location of colluvial landslides in a steep forested landscape' *IAHS-AISH publication*, 165, p.39-48.

Ricci, V. (2005) 'Fitting Distributions with R' [online] Available at: http://cran.r-project.org/doc/contrib/Ricci-distributions-en.pdf [Accessed 01/05/2012].

Rickli, C., and Graf, F. (2009) 'Effects of forests on shallow landslides–case studies in Switzerland' *Forest, Snow and Landscape Research*, *82***(1)**, p.33-44.

Robey, R. R. and Barcikowski, R. S. (1992) 'Type I Error and the Number of Iterations in Monte Carlo Studies of Robustness' *British Journal of Mathematical and Statistical Psychology,* 45**(2)**, p.283-288.

Roctek International (n/d) 'Slope Stabilisation Made Easier' [online] Available at: http://www.rotecinternational-usa.com/slopestabilization.htm [Accessed 06/03/2012].

Rosenblatt, M. (1956) 'Remarks on Some Nonparametric Estimates of a Density Function' *The Annals of Mathematical Statistics,* 27**(3)**, p.832-837.

Rossi, M., and Malamud, B. D. (2014) 'LAMPRE Report on Prototype Software for Landslide Statistics' *EU FP7 LAMPRE Project* [online] Available at: http://www.lampre-project.eu/index.php?option=com_phocadownload&view=category&id=26:wp5-triggered-event-landslides&Itemid=525 [Accessed 27/04/2015].

Rossi, M., Cardinali, M., Fiorucci, F., Marchesini, I., Mondini, A. C., Santangelo, M., and Guzzetti, F. (2012) 'A tool for the estimation of the distribution of landslide area in R' (EGU2012-9438) Presented at the European Geosciences Union General Assembly 2012, 22-27 April, Vienna, Austria.

Rossi, M., Guzzetti, F., Reichenbach, P., Mondini, A. C. and Peruccacci, S. (2010a) 'Optimal Landslide Susceptibility Zonation Based on Multiple Forecasts' *Geomorphology,* 114**(3)**, p.129-142.

Rossi, M., Witt, A., Guzzetti, F., Malamud, B. D. and Peruccacci, S. (2010b) 'Analysis of Historical Landslide Time Series in the Emilia-Romagna Region, Northern Italy' *Earth Surface Processes and Landforms,* 35**(10)**, p.1123-1137.

Sachs, M.K., Yoder, M.R., Turcotte, D.L., Rundle, J.B., Malamud, B.D. (2012) 'Black swans, power laws, and dragon-kings: Earthquakes, volcanic eruptions, landslides, wildfires, floods, and SOC models' *European Physical Journal,* 205, p.167-182.

Salvati, P., Balducci, V., Bianchi, C., Guzzetti, F. and Tonelli, G. (2009) 'A WebGIS for the Dissemination of Information on Historical Landslides and Floods in Umbria, Italy' *GeoInformatica,* 13**(3)**, p.305-322.

Santangelo, M., Cardinali, M., Rossi, M., Mondini, A. C., and Guzzetti, F. (2010) 'Remote landslide mapping using a laser rangefinder binocular and GPS' *Natural Hazards and Earth System Science*, *10***(12)**, p.2539-2546.

Scaioni, M., Longoni, L., Melillo, V. and Papini, M. (2014) 'Remote Sensing for Landslide Investigations: An Overview of Recent Achievements and Perspectives' *Remote Sensing,* 6**(10)**, p.9600-9652.

Scawthorn, C., Flores, P., Blais, N., Seligson, H., Tate, E., Chang, S., Mifflin, E., Thomas, W., Murphy, J. and Jones, C. (2006) 'HAZUS-MH Flood Loss Estimation Methodology. Ii. Damage and Loss Assessment' *Natural Hazards Review,* 7**(2)**, p.72-81.

Schneider, P.J. and Schauer, B. A. (2006) 'HAZUS—Its Development and Its Future' *Natural Hazards Review,* 7**(2)**, p.40-44.

Schulz, W. H., Galloway, S. L., and Higgins, J. D. (2012) 'Evidence for earthquake triggering of large landslides in coastal Oregon, USA' *Geomorphology*, 141, p.88-98.

Schuster, R. L. and Highland, L. M. (2003) 'Impact of Landslides and Innovative Landslide-Mitigation Measures on the Natural Environment' *In: Geologic Hazards Team, US Geological Survey, Denver, Colorado.*

Schuster, Robert L., and William J. Kockelman. 'Principles of landslide hazard reduction.' In Turner, K. and Schuster, R. (eds.) *Landslides Investigation and Mitigation Transportation Research Board Special Report 247*, Washington DC: Transportation Research Board, p.91-105.

Shahabi, H., Ahmad, B. B., and Khezri, S. (2013) 'Evaluation and comparison of bivariate and multivariate statistical methods for landslide susceptibility mapping (case study: Zab basin)' *Arabian Journal of Geosciences*, 6**(10)**, p.3885-3907.

Shapiro, M. and Waupotitsch, O. (2011) 'Grass GIS Manual: R.Slope.Aspect' [online] Available at: http://grass.osgeo.org/grass65/manuals/r.slope.aspect.html [Accessed 27/04/2015].

Shiraki, N., Shinozuka, M., Moore, J. E., Chang, S. E., Kameda, H. and Tanaka, S. (2007) 'System Risk Curves: Probabilistic Performance Scenarios for Highway Networks Subject to Earthquake Damage' *Journal of Infrastructure Systems,* 13**(1)**, p.43-54.

Sidle, R. C., and Ochiai, H. (2006) *Landslides: processes, prediction, and land use* (Vol. 18) American Geophysical Union.

Singhroy, V. (1995) 'SAR Integrated Techniques for Geohazard Assessment' *Advances in Space Research,* 15**(11)**, p.67-78.

Slice, D. E. (2007) 'Geometric Morphometrics' *Annual Review of Anthropology,* 36**(1)**, p.261-281.

Smith, D. (1994) 'Flood Damage Estimation- a Review of Urban Stage-Damage Curves and Loss Functions' *Water S. A.,* 20**(3)**, p.231-238.

Soeters, R., van Westen, C. (1996) 'Slope Instability Recognition, Analysis and Zonation' *In:* Turner, K. and Schuster, R. (eds.) *Landslides Investigation and Mitigation Transportation Research Board Special Report 247*, Washington DC: Transportation Research Board, p.129-177

Sohn, J. (2006) 'Evaluating the Significance of Highway Network Links under the Flood Damage: An Accessibility Approach' *Transportation Research Part A: Policy and Practice,* 40**(6)**, p.491-506.

Sökefeld, M. (2012) The Attabad landslide and the politics of disaster in Gojal, Gilgit-Baltistan. *Negotiating Disasters: Politics, Representations, Meanings*. Frankfurt: Peter Lang, p.175-204.

Soothill, K. and Grover, C. (1997) 'A Note on Computer Searches of Newspapers' *Sociology,* 31**(3)**, p.591-596.

Soroka, S. N., Farnsworth, S. J., Young, L. and Lawlor, A. (2009) 'Environment and Energy Policy: Comparing Reports from Us and Canadian Television News' *In: Annual Meeting of the American Political Science Association, Toronto, Canada, September*, 3-6.

Sprenger, T. O. and Welpe, I. M. (2011) 'News or Noise? The Stock Market Reaction to Different Types of Company-Specific News Events' [online] Available at: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1734632 [Accessed 27/04/215].

Stark, C. and Guzzetti, F. (2009) 'Landslide Rupture and the Probability Distribution of Mobilized Debris Volumes' *Journal of Geophysical Research: Earth Surface,* 114**(F2)**.

Stark, C. P.and Hovius, N. (2001) 'The Characterization of Landslide Size Distributions' *Geophys. Res. Lett.,* 28**(6)**, p.1091-1094.

Stillinger, F. H. and Weber, T. A. (1984) 'Inherent Pair Correlation in Simple Liquids' *The Journal of Chemical Physics,* 80**(9)**, p.4434-4437.

Stoyan, D. and Stoyan, H. (1996) 'Estimating Pair Correlation Functions of Planar Cluster Processes' *Biometrical Journal,* 38**(3)**, p.259-271.

Stucchi, M., Albini, P., Mirto, M. and Rebez, A. (2004) 'Assessing the Completeness of Italian Historical Earthquake Data' *Annals of Geophysics,* 47**(2-3)**

Sugai, T., Ohmori, H. and Hirano, M. (1995) 'Rock Control on Magnitude-Frequency Distribution of Landslide' *International Journal of Rock Mechanics and Mining Sciences and Geomechanics Abstracts,* 32**(3)**, p.116A-116A.

Supriana, I. and Aryan, P.R. (2011) 'Direct Skeleton Extraction Using River-Lake Algorithm' In: *Proceedings of International Conference on Electrical Engineering and Informatics (ICEEI), 2011*, 17-19 July 2011, Bandung, Indonesia, p.1-3.

Süzen, M (2002) *Data Driven Landslide Hazard Assessment Using Geographical Information Systems and Remote Sensing*, Ph.D. Thesis, Middle East Technical University.

TableCurve2D (2015) 'TableCurve2D Statistical Curve Fitting Software' [online] Available at: http://www.sigmaplot.co.uk/products/tablecurve2d/tablecurve2d.php [Accessed 27/04/2015].

Tan, Y. C., and Chow, C. M. (2004) 'Slope stabilization using soil nails: design assumptions and construction realities' *Presented at Malaysia-Japan Symposium on Geohazards and Geoenvironmental Engineering, Bangi, Malaysia* , December 2004, p.13-14.

Tarhule, A. (2005) 'Damaging rainfall and flooding: the other Sahel hazards' *Climatic change*, 72**(3)**, p.355-377.

Taylor, F. E., Malamud, B. D., Freeborough, K., Demeritt, D. (forthcoming) 'Enriching Great Britain's National Landslide Database by Searching Newspaper Archives' *Geomorphology*.

The Grand Alpine Tour. (2014) 'Grand Alpine Tour, Recipients of the RGS Land Rover Bursary, 2014' [online] Available at: http://grandalpinetour.com/ [Accessed 27/04/2015].

Tonini, M., Pedrazzini, A., Penna, I. and Jaboyedoff, M. (2014) 'Spatial Pattern of Landslides in Swiss Rhone Valley' *Natural Hazards,* 73**(1)**, p.97-110.

Transport Scotland (2012) 'Wig Wag Road Signs' [online] Available at: http://www.transport scotland.gov.uk/road/maintenance/prioritising-and-maintaining/landslides/wigwag [Accessed 06/03/2012].

Trauth, M. H., Bookhagen, B., Marwan, N. and Strecker, M. R. (2003) 'Multiple Landslide Clusters Record Quaternary Climate Changes in the Northwestern Argentine Andes' *Palaeogeography, Palaeoclimatology, Palaeoecology,* 194**(1)**, p.109-121.

Trimble, S. W. (2008) 'The Use of Historical Data and Artifacts in Geomorphology' *Progress in Physical Geography,* 32**(1)**, p.3-29.

Trustrum, N. (1988) 'Photograph of landslides caused by Cyclone Bola, 1088 in Napier, New Zealand' [online] Available at: http://www.teara.govt.nz/en/photograph/14550/ landslides-in-cleared-country [Accessed 27/04/2015].

Tschoegl, L., Below, R. and Guha-Sapir, D. (2006) 'An Analytical Review of Selected Data Sets on Natural Disasters and Impacts' *In: Centre for Research on the Epidemiology of Disasters.* Bangkok- Thailand.

Turcotte, D. L. (2006) 'Modeling Geocomplexity: "A New Kind of Science"' *Geological Society of America Special Papers,* 413, p.39-50.

Turcotte, D. L. and Malamud, B. D. (2004) 'Landslides, Forest Fires, and Earthquakes: Examples of Self-Organized Critical Behavior' *Physica A: Statistical Mechanics and its Applications,* 340**(4)**, p.580-589.

Turcotte, D. L., Malamud, B. D., Guzzetti, F. and Reichenbach, P . (2002) 'Self-Organization, the Cascade Model, and Natural Hazards' *Proceedings of the National Academy of Sciences of the United States of America,* 99**(Suppl 1)**, p.2530-2537.

Turcotte, D. L., Morein, G., Roberts, D. and Malamud, B. (1999) 'Catastrophic Resurfacing and Episodic Subduction on Venus' *Icarus,* 139**(1)**, p.49-54.

Turner, K. and Schuster, R. (eds.) *Landslides Investigation and Mitigation Transportation Research Board Special Report 247*, Washington DC: Transportation Research Board, p.439-473.

Uchida, H., and Nelson, A. (2009) 'Agglomeration index: towards a new measure of urban concentration' No. 2010, 29. Working paper//World Institute for Development. Economics Research, 2010.

UN (2014) *2014 Human Development Report* , New York: United Nations Development Programme.

UNISDR (2015) 'Sendai Framework for Disaster Risk Reduction 2015-2030' [online] Available at: http://www.wcdrr.org/uploads/Sendai_Framework_for_Disaster_Risk_Reduction _2015-2030.pdf [Accessed 27/04/2015].

USCB (2012) 'Tiger Products: Tiger/Line Shapefiles & Files' *United States Census Bureau* [online] Available at: http://www.census.gov/geo/www/tiger/ [Accessed 01/05/2012].

USGS (2006) 'SRTM Digital Elevation Model' [online] Available at: http://srtm.usgs.gov/ [Accessed 27/04/2015].

USGS (2015) 'Earthquake Hazards Program' [online] Available at: http://earthquake.usgs.gov/ [Accessed 27/04/2015].

USGS (2005) 'Preliminary Integrated Geologic Map Databases for the United States - Western States: California, Nevada, Arizona, Washington, Oregon, Idaho, and Utah' Open File Report 2005-1305, United States Geological Survey.

USGS (2013) 'Landslide Photo Collections' [online] Available at: http://landslides.usgs.gov/learn/photos/ [Accessed 27/04/2015].

van Asch, T. W. J., Malet, J.-P., van Beek, L. P.H. and Amitrano, D. (2007) 'Techniques, Issues and Advances in Numerical Modelling of Landslide Hazard' *Bulletin de la Societe Geologique de France,* 178**(2)**, p.65-88.

van Atteveldt, W., Kleinnijenhuis, J. and Ruigrok, N. (2008) 'Parsing, Semantic Networks, and Political Authority Using Syntactic Analysis to Extract Semantic Relations from Dutch Newspaper Articles' *Political Analysis,* 16**(4)**, p.428-446.

Van Den Eeckhaut, M. and Hervás, J. (2012) 'State of the Art of National Landslide Databases in Europe and Their Potential for Assessing Landslide Susceptibility, Hazard and Risk' *Geomorphology,* 139–140**(0)**, p.545-558.

Van Den Eeckhaut, M., Poesen, J., Govers, G., Verstraeten, G. and Demoulin, A. (2007) 'Characteristics of the Size Distribution of Recent and Historical Landslides in a Populated Hilly Region' *Earth and Planetary Science Letters,* 256**(3–4)**, p.588-603.

Van Leeuwen, J. M. J., Groeneveld, J., and De Boer, J. (1959) 'New method for the calculation of the pair correlation function' *Physica*, 25**(7)**, p.792-808.

Van Leeuwen, J., Groeneveld, J. and De Boer, J. (1959) 'New Method for the Calculation of the Pair Correlation Function. I' *Physica,* 25**(7)**, p.792-808.

van Westen, C. J., Castellanos, E. and Kuriakose, S. L. (2008) 'Spatial Data for Landslide Susceptibility, Hazard, and Vulnerability Assessment: An Overview' *Engineering Geology,* 102**(3–4)**, p.112-131.

van Westen, C. J., Rengers, N., Soeters, R. (2003) 'Use of Geomorphological Information in Indirect Landslide Susceptibility Assessment' *Natural Hazards*, 30, p.399-419.

van Westen, C. J., van Asch, T. W. J. and Soeters, R. (2006) 'Landslide Hazard and Risk Zonation—Why Is It Still So Difficult?' *Bulletin of Engineering Geology and the Environment,* 65**(2)**, p.167-184.

Varnes, D. J. (1978) Slope Movements Types and Processes. *In, Special Report 176: Landslides: Analysis and Control.* Washington DC: Transportation Research Board.

Voumard, J., Caspar, O., Derron, M. H., and Jaboyedoff, M. (2013) 'Dynamic risk simulation to assess natural hazards risk along roads' *Natural Hazards and Earth System Science*, 13**(11),** p.2763-2777.

Wang, H. B., Sassa, K. and Xu, W. Y. (2007) 'Analysis of a Spatial Distribution of Landslides Triggered by the 2004 Chuetsu Earthquakes of Niigata Prefecture, Japan' *Natural Hazards,* 41**(1)**, p.43-60.

Ward-Perkins, J. B. (1962) 'Etruscan Towns, Roman Roads and Medieval Villages: The Historical Geography of Southern Etruria' *The Geographical Journal,* 128**(4)**, p.389-404.

Washington Division of Geology and Earth Resources. (2014) 'Washington State Landslides and Landforms--GIS Data, Washington Division of Geology and Earth Resources Digital Data Series Ds-12, Version 2.0.' [online] Available at: https://fortress.wa.gov/dnr/geology/?Theme=natural_hazards [Accessed 27/04/2015].

Watson, C. C. and Johnson, M. E. (2004) 'Hurricane Loss Estimation Models: Opportunities for Improving the State of the Art' *Bulletin of the American Meteorological Society,* 85**(11)**, p.1713-1726.

Watts, D. J. and Strogatz, S. H. (1998) 'Collective Dynamics of 'Small-World' Networks' *Nature,* 393**(6684)**, p.440-442.

Weaver, D. A. and Bimber, B. (2008) 'Finding News Stories: A Comparison of Searches Using Lexisnexis and Google News' *Journalism & Mass Communication Quarterly,* 85**(3)**, p.515-530.

Wen, Q., He, H., Wang, X., Wu, W., Wang, L., Xu, F., Wang, P., Tang, T. and Lei, Y. (2011) 'UAV Remote Sensing Hazard Assessment in Zhouqu Debris Flow Disaster' *SPIE Remote Sensing,* p.817510-817510-8.

West, D. B. (2001) *Introduction to Graph Theory*. Upper Saddle River : Prentice Hall.

White, E. P., Enquist, B. J. and Green, J. L. (2008) 'On Estimating the Exponent of Power-Law Frequency Distributions' *Ecology,* 89**(4)**, p.905-912.

White, J. (2010) 'Doing Maximum Likelihood Estimation by Hand in R' [online] Available at: http://www.johnmyleswhite.com/notebook/2010/04/21/doing-maximum-likelihood-estimation-by-hand-in-r/ [Accessed 27/04/2015].

Whitehouse, I.E., Griffiths, G. A. (1983) 'Frequency and hazard of large rock avalanches in the Central Southern Alps, New Zealand' *Geology*, 11p.331–334.

Wilcox, R. (2005) 'Kolmogorov–Smirnov Test' In: Armitage, P., Colton. C. (eds.) *Encyclopaedia of Biostatistics*. Chichester: John Wiley & Sons, Ltd.

Wills, C., J. and McCrink, T. P . (2002) 'Comparing Landslide Inventories: The Map Depends on the Method' *Environmental & Engineering Geoscience,* 8**(4)**, p.279-293.

Wills, C., Perez, F., and Branum, D. (2014) 'New Method for Estimating Landslide Losses from Major Winter Storms in California and Application to the ARkStorm Scenario' *Natural Hazards Review, doi:10.1061/(ASCE)NH.1527-6996.0000142 , A4014001.*

Winston, W. L. (1996) *Simulation Modelling Using @Risk*. Pacific Grove: Duxbury Press.

Winter, M. G. (2015) 'The Vulnerability Shadow Cast by Debris Flow Events' In: Lollino, G., Giordan, D., Thuro, K., Carranza-Torres, C., Wu, F., Marinos P., and Delgado C. (eds.) *Engineering Geology for Society and Territory - Volume 6*. Switzerland: Springer International Publishing, p.641-644.

Winter, M. G., Macgregor, F., Shackman, L. (eds) (2008) *Scottish Road Network Landslides Study: Implementation*, Edinburgh: Transport Scotland.

Winter, M., Macgregor, F. and Shackman, L. (2005) *Scottish Road Network Landslides Study Summary Report*, Edinburgh: Transport Scotland.

Woolley, R. R., Marsell, R. E. and Grover, N. C. (1946) *Cloudburst Floods in Utah, 1850-1938*. US Government Printing Office.

World Bank (2014) 'New Country Classifications of Economic Status' [online] Available at: http://data.worldbank.org/news/new-country-classifications [Accessed 27/04/2015].

WSL (Swiss Federal Institute for Forest, Snow and Landscape Research) (2011) 'Flexible Rockfall Protection Gallery Systems' [online] Available at: http://www.wsl.ch/fe /gebirgshydrologie/massenbewegungen/projekte/KTI_Galleries/index_EN [Accessed 06/03/2012].

Xie, F. and Levinson, D. (2007) 'Measuring the Structure of Road Networks' *Geographical Analysis,* 39**(3)**, p.336-356.

Xu, C. (2014) 'Preparation of earthquake-triggered landslide inventory maps using remote sensing and GIS technologies: Principles and case studies' *Geoscience Frontiers*, Advanced article in press, doi:10.1016/j.gsf.2014.03.004.

Xu, C. and Xu, X. (2012) 'Comment on "Spatial Distribution Analysis of Landslides Triggered by 2008.5. 12 Wenchuan Earthquake, China" by Shengwen Qi, Qiang Xu, Hengxing Lan, Bing Zhang, Jianyou Liu [Engineering Geology 116 (2010) 95–108].' *Engineering Geology,* 133, p.40-42.

Xu, C., Shyu, J. B. H. and Xu, X. (2014) 'Landslides Triggered by the 12 January 2010 Port-Au-Prince, Haiti, Mw = 7.0 Earthquake: Visual Interpretation, Inventory Compiling, and Spatial Distribution Statistical Analysis' *Nat. Hazards Earth Syst. Sci.,* 14**(7)**, p.1789-1818.

Xu, C., Xu, X., Dai, F., Xiao, J., Tan, X., and Yuan, R. (2012) 'Landslide hazard mapping using GIS and weight of evidence model in Qingshui river watershed of 2008 Wenchuan earthquake struck region' *Journal of Earth Science*, 23, p.97-120.

Yagi, H., Sato, G., Higaki, D., Yamamoto, M., and Yamasaki, T. (2009) 'Distribution and characteristics of landslides induced by the Iwate–Miyagi Nairiku Earthquake in 2008 in Tohoku District, Northeast Japan' *Landslides*,6**(4)**, p.335-344.

Yang, W., Shen, L. and Shi, P.2015. Mapping Landslide Risk of the World. *In,* eds.) (Shi and Kasperson)*World Atlas of Natural Disaster Risk*. Springer Berlin Heidelberg. p.57-66

Yang, Z., Lee, Y. (2006) 'The Fractal Characteristics of Landslides Induced by Earthquakes and Rainfall in Central Taiwan' *In: Proceedings of the 10th International Congress of the International Association for Engineering Geology and the Environment, Nottingham, 6-10th September 2006*, Bedrock, p.48-56.

Yin, Y., Wang, F. and Sun, P . (2009) 'Landslide Hazards Triggered by the 2008 Wenchuan Earthquake, Sichuan, China' *Landslides,* 6**(2)**, p.139-152.

Yuill, R. S. (1971) 'The Standard Deviational Ellipse; an Updated Tool for Spatial Description' *Geografiska Annaler. Series B, Human Geography,* 53**(1)**, p.28-39.

Zaitchik, B. F., van Es, H. M., and Sullivan, P.J. (2003) 'Modelling Slope Stability in Honduras' *Soil Science Society of America Journal*, 67**(1)**, p.268-278.

Zare, M., Pourghasemi, H. R., Vafakhah, M., and Pradhan, B. (2013) 'Landslide susceptibility mapping at Vaz Watershed (Iran) using an artificial neural network model: a

comparison between multilayer perceptron (MLP) and radial basic function (RBF) algorithms' *Arabian Journal of Geosciences*, 6**(8)**, p.2873-2888.

Zêzere, J., Oliveira, S., Garcia, R. and Reis, E. (2007) 'Landslide Risk Analysis in the Area North of Lisbon (Portugal): Evaluation of Direct and Indirect Costs Resulting from a Motorway Disruption by Slope Movements' *Landslides,* 4**(2)**, p.123-136.

Zhou, W., Beck, B. and Adams, A. (2003) 'Application of Matrix Analysis in Delineating Sinkhole Risk Areas along Highway I-70, near Frederick, Maryland' *Environmental Geology,* 44**(7)**, p.834-842.

Appendix A: Python and R Codes to Analyse Landslide Shape

This appendix A contains:

A.1 **Python code** (executable in Grass-GIS) which fits elliptical shapes to real landslide polygons and tests the goodness of fit using Lombardo (2014)'s measure of ellipticity (e). The user must input a .csv file containing the centroid (x,y) locations, orientation, major and minor axes of the ellipses to be fitted (the method to create these values is outlined in **Chapter 5**) and a shapefile of the observed triggered landslide event inventory they wish to fit the ellipses to. The code then:

(i) Plots the landslide ellipses
(ii) Overlays these ellipses with the original landslide polygons to calculate the area of intersection (A_I)
(iii) Calculates ellipticity (e) based on A_I and the original landslide polygon area (A_L)
(iv) Outputs shapefiles of an inventory of A_I and landslide elliptical approximations

A.2 **R code** (executable in R statistical software) to analyse landslide L/W statistical distribution (the output of code **A.1)**. Before this script is run, the user must filter out any landslides where ellipticity (e) < 0.5 (or whichever threshold they select) and separate landslides into categories of landslide area (A_L). The filtered, binned values of L/W are input to R as a *.csv file. For each landslide area category, the code performs:

(i) Basic frequency-size analysis (and plotting)
(ii) Maximum likelihood estimation to fit the inverse gamma probability density function to landslide L/W
(iii) Monte Carlo Kolmogorov-Smirnov goodness of fit testing
(iv) Bootstrapping to estimate parameter uncertainty
(v) Summary plots, output as a *.pdf file

Appendix A.1 Python Grass GIS Code to Calculate Landslide Ellipses

```
#!/usr/bin/env python

############################################################################################
#                                                                                          #
# MODULE:       v.landslide_ellipses.py                                                    #
#                                                                                          #
# AUTHOR(S):    Faith Taylor                                                               #
# PURPOSE:      Create elliptical approximations of landslide shapes and measure their     #
               ellipticity goodness of fit                                                #
# DATE:         1st April 2015                                                             #
# COPYRIGHT:(C) 2015  Faith Taylor                                                         #
#                                                                                          #
#         This program is free software under the GNU General Public                       #
#         License (>=v2). Read the file COPYING that comes with GRASS                      #
```

```
#         for details.                                                                    #
#                                                                                         #
###########################################################################################

#%module
#% description: Create elliptical approximations of landslide shapes and measure their
ellipticity goodness of fit. The user must provide a csv file containing: Landslide ID, x,y
locations, landslide area, ellipse long and short axis, ellipse orientation. See manual for
more information how to create this
#% keywords: vector
#% keywords: landslide
#%end
#%option
#% key: true_ls
#% type: string
#% gisprompt: old,vector,vector
#% description: The real landslide inventory
#% required: yes
#%end
#%option
#% key: inventory_type
#% type: string
#% description: The type of shapes we are fitting to the landslides. Creates a suffix to
each vector filename
#% required: yes
#%end
#%option
#% key: filepath
#% type: string
#% required: yes
#% multiple: no
#% key_desc: name
#% label: Name of input file to be imported containing landslide ellipse info to be plotted
#% description: '-' for standard input
#% gisprompt: old,file,file
#%end

# Import necessary modules
from __future__ import division
import math
from numpy import *
import shlex
import sys
import os
import atexit
import grass.script as grass
from grass.script import core as grasscore
def cleanup():
        pass

def main():
        # Assign input files a variable name
        true_ls = options['true_ls']
        inventory_type = options['inventory_type']
        filepath = options['filepath']

        #create column headers
        columns = ['orig_ls_id integer',
                        'east double precision',
                        'north double precision',
                        'radiusL double precision',
                        'radiusW double precision',
                        'azimuth double precision',
                        'area_ls double precision',
                        'perim_ls double precision']

        # Create a vector map of randomly generated points. The accompanying table contains
        a landslide area and length to width ratio for each point.
        grass.run_command('v.in.ascii', cat = 0, x = 2, y = 3, separator = 'comma',
                        input = filepath,
                        output = "real_ls_centroids",
```

```
                columns = columns,
                skip = 1,
                overwrite = True)

grass.run_command('v.db.addcolumn',
                map = "real_ls_centroids",
                columns = "area_elip double precision, area_elip_clip double
                precision, ellipticity double precision")

#Read category values of each point mapped
category_vals = grass.read_command("v.category",
                input = "real_ls_centroids",
                type = "point",
                option = "print",
                flags = "g")

category_vals_a = category_vals.split('\n')
num_points = len(category_vals_a)
patch_list = []

# Start a loop to buffer each point with an ellipitcal shape and overlay with
original landslide. THis is done one by one in case any original landslides are
overlapping, which creates topology issues
for i in range(0, num_points-1):
        category_value = category_vals_a[i]

        #Extract the individual point, which will become the landslide centroid
        grass.run_command("v.extract",
                input = "real_ls_centroids",
                output = "point"+str(category_value),
                cats = category_value,
                overwrite = True)

        #Read the major & minor axes and the rotation (aspect) of the landslide from
        the point file
        major = grass.read_command("v.db.select",
                flags = "c",
                map = "point"+str(category_value),
                columns = "radiusL")

        minor = grass.read_command("v.db.select",
                flags = "c",
                map = "point"+str(category_value),
                columns = "radiusW")


        aspect = grass.read_command("v.db.select",
                        flags = "c",
                        map = "point"+str(category_value),
                        columns = "azimuth")

        #Add a buffer around that point to give the landslide an area
        grass.run_command("v.buffer",
                                overwrite = True,
                                input = "point"+str(category_value),
                                output = str(inventory_type) +"Ellipse_landslide"
                                +str(category_value),
                                distance = major,
                                minordistance = minor,
                                angle = aspect,
                                flags = "t")

        # Verify the area of this ellipse
        grass.run_command("v.to.db",
                        map = str(inventory_type) +"Ellipse_landslide"
                        +str(category_value),
                        option = "area",
                        columns = "area_elip")

        #Extract the matching real landslide
        grass.run_command("v.extract",
```

```
                        input = true_ls,
                        output = "Real_landslide"+str(category_value),
                        cats = str(category_value),
                        overwrite = True)

        #Overlay this elliptical approximation of the landslide with the original
        landslide shape, where the category values should match
        grass.run_command("v.overlay",
                                overwrite = True,
                                ainput = "Real_landslide"+str(category_value),
                                binput = str(inventory_type)
                                +"Ellipse_landslide"+str(category_value),
                                operator = "and",
                                output = str(inventory_type)
                                +"AE_clip"+str(category_value))

        # Upload the area of this overlap to the file
        grass.run_command("v.to.db",
                        map = str(inventory_type) +"AE_clip"+str(category_value),
                        option = "area",
                        columns = "b_area_elip_clip")

        #Calculate the ellipticity of original shape
        grass.run_command("v.db.update",
                map = str(inventory_type) +"AE_clip"+str(category_value),
                column = "b_ellipticity",
                value = "1-((2*(b_area_ls - b_area_elip_clip))/b_area_ls)")

# Patch all individual landslide ellipse vector files into one map
patch_list = []
for f in range (1, num_points):
        patch_name = (str(inventory_type)+'Ellipse_landslide'+str(f)+",")
        patch_list.append(patch_name)
len_patch_list = len(patch_list)
num_patches = int((len_patch_list/100))
super_patch_list = []
for i in range(0, num_patches):
        patch_list_a  = patch_list[(i*100):((i*100)+100)]
        patch_list_2 = ''.join(patch_list_a)
        if len(patch_list_a) > 1:
                grass.run_command("v.patch",
                                        input = ( patch_list_2),
                                        output = "inventory_patch"+str(i),
                                        flags = "e",
                                        overwrite = True)
                super_patch_list.append("inventory_patch"+str(i)+",")
        else:
                super_patch_list.append(patch_list_2)

super_patch_list_2 = []

for k in range(1, len(super_patch_list)):
        ind_patch_name = ("inventory_patch"+str(k)+",")
        super_patch_list_2.append(ind_patch_name)

grass.run_command("v.patch",
                        input = ( super_patch_list_2),
                        output = str(inventory_type) +"Ellipse_inventory_final",
                        flags = "e",
                        overwrite = True)

# Patch all individual landslide ellipse/original landslide intersection area vector
files into one map
patch_list = []
for f in range (1, num_points):
        patch_name = (str(inventory_type)+'AE_clip'+str(f)+",")
        patch_list.append(patch_name)


len_patch_list = len(patch_list)
num_patches = int((len_patch_list/100))
```

```
        super_patch_list = []

        for i in range(0, num_patches+1):
                patch_list_a  = patch_list[(i*100):((i*100)+100)]
                patch_list_2 = ''.join(patch_list_a)

                if len(patch_list_a) > 1:
                        grass.run_command("v.patch",
                                          input = ( patch_list_2),
                                          output = "clip_inventory_patch"+str(i),
                                          flags = "e",
                                          overwrite = True)

                        super_patch_list.append("clip_inventory_patch"+str(i)+",")
                else:
                        super_patch_list.append(patch_list_2)

        super_patch_list_2 = []

        for k in range(1, int((math.ceil((num_patches/100))))):
                ind_patch_name = ("clip_inventory_patch"+str(k)+",")
                super_patch_list_2.append(ind_patch_name)

        grass.run_command("v.patch",
                                input = ( super_patch_list_2),
                                output = str(inventory_type)+"Clip_inventory_final",
                                flags = "e",
                                overwrite = True)

if __name__ == "__main__":
        options, flags = grass.parser()
        atexit.register(cleanup)
        sys.exit(main())
```

Appendix A.2 R Code to Analyse the Statistical Distribution of Landslide Length to Width Ratio

```
#############################################################################################
#                                                                                           #
# Name: Inverse Gamma Landslide Fit                                                         #
# Author: Faith Taylor                                                                      #
# Last updated : 1st April 2015                                                             #
# Purpose: Fit inverse gamma pdf to landslide data and test goodness of fit                 #
# Copyright: (c) 2014 Faith Taylor                                                          #
# This program is free software under the GNU General Public Licence (>=v2).                #
#                                                                                           #
#############################################################################################

# Load modules
library(MCMCpack)
library(MASS)
library(PearsonDS)
library(calibrate)
library(gplots)

# 1. Initial definitions
# 1.1 The inverse gamma log likelihood function
ll = function(par){
  if(par[1]>0 & par[2]>0 & par[3]<min(data)) return( -sum(log(dinvgamma(data-
par[3],par[1],par[2]))) )
  else return(Inf)
}
```

```
# 1.2 Set place name and output file
Location = "taiwan"
location = "Taiwan"
pdf_name = paste(location,"_lwr_analysis.pdf",sep="")
pdf(paste(pdf_name),width = 11, height = 9)

# 1.3 Read Data file. L/W values must be pre-binned into area categories
alldata = paste(Location,"_lwr_by_area", sep = "")
alldata = get(alldata)

# 1.4 Set the text for the area category labels
lower = lower_upper_vals[1,2]
upper = lower_upper_vals[1,3]
phrase1 = bquote(bold(.(lower)~ m^2 ~""<= "" ~bolditalic(A[L]) ~ "<" ~bold(.(upper) ~ m^2)))
lower = lower_upper_vals[2,2]
upper = lower_upper_vals[2,3]
phrase2 = bquote(bold(.(lower)~ m^2 ~""<= "" ~bolditalic(A[L]) ~ "<" ~bold(.(upper) ~ m^2)))
lower = lower_upper_vals[3,2]
upper = lower_upper_vals[3,3]
phrase3 = bquote(bold(.(lower)~ m^2 ~""<= "" ~bolditalic(A[L]) ~ "<" ~bold(.(upper) ~ m^2)))
lower = lower_upper_vals[4,2]
upper = lower_upper_vals[4,3]
phrase4 = bquote(bold(.(lower)~ m^2 ~""<= "" ~bolditalic(A[L]) ~ "<" ~bold(.(upper) ~ m^2)))
lower = lower_upper_vals[5,2]
upper = lower_upper_vals[5,3]
phrase5 = bquote(bold(.(lower)~ m^2 ~""<= "" ~bolditalic(A[L]) ~ "<" ~bold(.(upper) ~ m^2)))
lower = lower_upper_vals[6,2]
upper = lower_upper_vals[6,3]
phrase6 = bquote(bold(.(lower)~ m^2 ~""<= "" ~bolditalic(A[L]) ~ "<" ~bold(.(upper) ~ m^2)))
lower = lower_upper_vals[7,2]
upper = lower_upper_vals[7,3]
phrase7 = bquote(bold(.(lower)~ m^2 ~""<= "" ~bolditalic(A[L]) ~ "<" ~bold(.(upper) ~ m^2)))
lower = lower_upper_vals[8,2]
upper = lower_upper_vals[8,3]
phrase8 = bquote(bold(.(lower)~ m^2 ~""<= "" ~bolditalic(A[L]) ~ "<" ~bold(.(upper) ~ m^2)))
lower = lower_upper_vals[9,2]
upper = lower_upper_vals[9,3]
phrase9 = bquote(bold(.(lower)~ m^2 ~ ""<= "" ~bolditalic(A[L]) ~ "<" ~bold(.(upper) ~m^2)))
lower = lower_upper_vals[10,2]
upper = lower_upper_vals[10,3]
phrase10 = bquote(bold( ~bolditalic(A[L]) ~ ""<= "" ~bold(.(lower) ~ m^2)))

# 1.5 Make a series of x values (Length to width ratios)
xs = seq(1,100, 0.01)

# 1.6 Set display parameters
par(mar= c(8,5,1,1), family = "serif")

# 1.7 Define function to round up to the nearest 1, 10, 100, 1000 etc.
roundUp <- function(x, nice=c(1,2,4,5,6,8,10)) {
  if(length(x) != 1) stop("'x' must be of length 1")
  10^floor(log10(x)) * nice[[which(x <= 10^floor(log10(x)) * nice)[[1]]]]
}

# 2. Start loop to analyse L/W values in each area category
for (A in 1:10){
  # Delete any variables from the previous loop
  rm(shape_f, location_f, scale_f, Proportion, all_d, quants, m, ests, estsa, b, bin,
  bin_boundaries, binsizes, bootstrap_location, bootstrap_location, bootstrap_scale,
  bootstrap_scale_recip, bootstrap_shape, cumulative_x, dif, data, dval, emp_bin_mids,
  emp_cum_y, emp_freq_dens, emp_freqs, emp_prob_dens, empirical_D, g, h, i, k, label,
  loc_max, location_s, logdat, loghist, lower, lower_pdf, max_pdf, max_x_d, maxx,
  median_pdf,  mle, n, numbins, params, parset, pdf,  rand_D, rand_cdf, rand_lwrs,
  rand_lwrs_cumulative, location_1, scale_a, shape_a, shape_rho, smooth_teo_cum,
  startlocation, startscale, startshape, teo_cdf, teo_cum, testresult, upper, upper_pdf,
  val, yupper, yupper2)

  # Set the label text for this specfic area category
  lower = lower_upper_vals[A,2]
  upper = lower_upper_vals[A,3]
  phrase = bquote(bold(.(lower)~ m^2 ~"<" ~bolditalic(A[L]) ~ "<" ~bold(.(upper) ~ m^2)))
```

```
  variablej = paste("phrase_", A, sep = "")
  assign(variablej, phrase)

  # Read data
  data = alldata[A]
  data = data[!is.na(data)]
  run = paste("LW", A, sep = "")
  all_data_with_no_NAs_but_no_titles = alldata[!is.na(alldata)]
  max_data = max(all_data_with_no_NAs_but_no_titles)
  round_max_data = roundUp(max_data)

  # Some inventories may not have any data in a particular area category, in which case skip
   this loop
  if ((length(data))<1){
    next
  } else {

  #Calculate the number of observations
  n = length(data)
  phrase_a = bquote(bolditalic( "n = "~.(n)))


 # Fitting an inverse gamma pdf to the raw data
  # Use mle fitting to find the parameters that best fit the lwrs in this area category.
  This uses the optim function
  try((mle = optim(c(1,1,1),ll)), silent = FALSE)

  # Obtain initial parameter values
  params = mle$par
  shape_rho = params [1]
  scale_a = params[2]
  location_s = params[3]

  # Produce a plot of the observed data and the pdf/cdf fit to the data (non cumulative then
  cumulative)
  # Evaluate the inverse gamma pdf for each of the x values using the parameters from MLE
  fitting
  pdf = dpearsonV(xs, shape_rho, location_s, (1/scale_a))

  # Bin the raw data
  logdat = log(data)
  loghist = hist(logdat, plot = FALSE)
  bin_boundaries = exp(loghist$breaks)
  emp_bin_mids = exp(loghist$mids)
  emp_freqs = loghist$counts
  #calculate bin sizes
  numbins = length(emp_bin_mids)
  binsizes = numeric(numbins)
  for (k in 1:numbins){
    dif = bin_boundaries[k +1] - bin_boundaries[k]
    binsizes[k] = dif}

  emp_freq_dens = emp_freqs/binsizes
  emp_prob_dens = emp_freqs/(binsizes * n)

  # Plot the pdf and raw data
  par(mar= c(5,6,2,2))
  plot(xs, pdf, type = "l", col = rgb(1,0,0), lwd = 3, ylim= c(10^-8, 10), log = "xy", ylab
= "", xlab=expression(bold("Length to Width Ratio, "~bolditalic(L/W))), axes = FALSE, yaxs
="i", xaxs = "i", cex.lab = 1.7 )
  ticks = c(0.1,1,2,3,4,5,6,7,8,9,10,20,30,40,50,60,70,80,90,100)
  title(ylab = expression(bold("Probability Density, " ~bolditalic(p(L/W)))), cex.lab = 1.7,
line = 4)
  axis(1, at = c(1,10, 100), labels =( expression(~bold(10^0),~bold(10^1), ~bold(10^2))),
lty = 1, lwd = 0, tck = 0, cex.axis = 1.7)
  yticks = c(0.000000001, 0.000000002, 0.000000003, 0.000000004, 0.000000005,  0.000000006,
0.000000007, 0.000000008, 0.000000009, 0.00000001, 0.00000002, 0.00000003, 0.00000004,
0.00000005,  0.00000006, 0.00000007, 0.00000008, 0.00000009,0.0000001, 0.0000002, 0.0000003,
0.0000004, 0.0000005,  0.0000006, 0.0000007, 0.0000008, 0.0000009, 0.000001, 0.000002,
0.000003,  0.000004, 0.000005, 0.000006, 0.000007, 0.000008, 0.000009, 0.00001, 0.00002,
0.00003, 0.00004, 0.00005, 0.00006, 0.00007, 0.00008, 0.00009,0.0001, 0.0002, 0.0003,
```

```
0.0004, 0.0005, 0.0006, 0.0007, 0.0008, 0.0009, 0.001, 0.002, 0.003, 0.004, 0.005, 0.006,
0.007, 0.008, 0.009, 0.01, 0.02, 0.03, 0.04, 0.05, 0.06, 0.07, 0.08, 0.09, 0.1, 0.2, 0.3,
0.4,0.5,0.6,0.7,0.8,0.9,1,2,3,4,5,6,7,8,9,10)
  axis(2, at = c(0.000000001, 0.00000001,0.0000001, 0.000001, 0.00001, 0.0001, 0.001, 0.01,
0.1, 1,10), labels =(expression(~bold(10^-9),~bold(10^-8),~bold(10^-7),~bold(10^-
6),~bold(10^-5), ~bold(10^-4), ~bold(10^-3), ~bold(10^-2), ~bold(10^-1), ~bold(10^0) ,
~bold(10^1))), lty = 1, lwd = 0, tck = 0, cex.axis = 1.7, las = 1)
  axis(1, ticks, labels = NA, lty = 1, lwd = 2, tck = 0.01)
  axis(2, yticks, labels = NA, lty = 1, lwd = 2, tck = 0.01)
  axis(3, ticks, labels = NA, lty = 1, lwd = 2, tck = 0.01)
  axis(4, yticks, labels = NA, lty = 1, lwd = 2, tck = 0.01)
  points(emp_bin_mids, emp_prob_dens, pch = 20)
  legend("topright", do.call("expression", list(location, phrase, phrase_a, "Observed",
"Inverse Gamma pdf")), lty = c(NA, NA, NA, NA, 1), col = c(NA, NA, NA, "black", "red"), pch
= c(NA, NA, NA, 20, NA), lwd = c(NA, NA, NA, NA, 2),bty = 'o', cex = 1.1, box.lwd = 2, inset
= 0.01)

  ## Now plot the cumulative distribution
  # Sort the observed data
  cumulative_x = sort(data)
  emp_cum_y = (1:n)/n
  emp_cum_y = sort(emp_cum_y, decreasing = TRUE)
  teo_cum = ppearsonV(cumulative_x, shape_rho, location_s, (1/scale_a), lower.tail = FALSE)
  empirical_D = max(abs(emp_cum_y - teo_cum))
  smooth_teo_cum = ppearsonV(xs, shape_rho, location_s, (1/scale_a), lower.tail = FALSE)

  par(mar= c(5,6,2,2))
  #Plot both distributions and check manually
  plot(xs, smooth_teo_cum, col = "red", type ="l", lwd = 2, axes = FALSE, xaxs = "i", yaxs =
"i", xlim = c(1,  ((max(data))+10)), ann = FALSE, ylim = c(0,1))
  points(cumulative_x, emp_cum_y, col = rgb(0,0,1,0.5), pch = 4, cex = 0.8)
  title(ylab = expression(bold("Cumulative Probability, " ~bolditalic(F(L/W)))), cex.lab =
1.5, line = 3)
  title(xlab = expression(bold("Length to Width Ratio, " ~bolditalic(L/W))), cex.lab = 1.5)
  axis(1, lwd = 2, tck =0.01, at = c(1,(seq(-1, ((max(data))+20), 1))), labels = FALSE)
  axis(3, lwd = 2, tck =0.01, at = c(1,(seq(-1, ((max(data))+20), 1))), labels = FALSE)
  axis(1, lwd = 2, tck =0.03, cex.axis = 1.5, font =2, at = c(1,(seq(0, ((max(data))+10),
5))))
  axis(2, lwd = 2, tck =0.03, cex.axis = 1.5, font =2, at = seq(0,1, by = 0.2), las =1)
  axis(2, lwd = 2, tck =0.01, cex.axis = 1.5, font =2, at = seq(0,1, by = 0.1), labels =
FALSE)
  axis(4, lwd = 2, tck =0.01, cex.axis = 1.5, font =2, at = seq(0,1, by = 0.1), labels =
FALSE)
  axis(3, lwd = 2, tck =0.03, labels = FALSE, at = c(1,(seq(0, ((max(data))+10), 5))))
  axis(4, lwd = 2, tck =0.03, labels = FALSE, at = seq(0,1, by = 0.2), las =1)
  legend("topright", do.call("expression", list(location, phrase, phrase_a, "Observed",
"Inverse Gamma cdf", phrase_a)), lty = c(NA,NA,NA, NA, 1), col = c(NA,NA,NA, "blue", "red"),
pch = c(NA,NA,NA, 20, NA), lwd = c(NA,NA,NA, NA, 2),bty = 'o', bg = "white", cex = 1.1,
box.lwd = 2, inset = 0.01)

  # Find the L/W value at which the maximum deviation between observed and fit cumulative
distribution occurs
  all_d = emp_cum_y - teo_cum
  g = match((min(emp_cum_y - teo_cum)), all_d)
  h = match((max(emp_cum_y - teo_cum)), all_d)
  if (abs(all_d[g]) > abs(all_d[h])) maxx = g else maxx = h
  max_x_d = cumulative_x[maxx]

  # Plot all D values
  par(mar= c(5,6,2,2))
  plot(cumulative_x, all_d, xlab = "L/W",  col ="darkgreen", pch = 20, cex = 0.5, xlim =
c(1, (max(data)+5)), ylim =c(-0.2, 0.2), xaxs ="i", yaxs = "i", ann = FALSE, axes = FALSE )
  abline(0,0, lty = 2, lwd = 2, col ="grey" )
  points(cumulative_x, all_d, xlab = "L/W",  col = rgb(0.7,1,0.7,0.5), type = "l")
  points(cumulative_x, all_d, xlab = "L/W",  col = "darkgreen", type = "p", pch = 20)
  roundy = roundUp(((max(data))+10))
  axis(1, lwd = 2, tck =0.02, cex.axis = 1.5, font =2, at = c(0,(seq(0, roundy, 5))))
  axis(1, lwd = 2, tck =0.01, cex.axis = 1.5, font =2, at = c(0,(seq(0, ((max(data))+10),
1))), labels = FALSE)
  axis(3, lwd = 2, tck =0.02, labels = FALSE, at = c(1,(seq(0, ((max(data))+10), 5))))
  axis(3, lwd = 2, tck =0.01, cex.axis = 1.5, font =2, at = c(1,(seq(0, ((max(data))+10),
```

```
1))), labels = FALSE)
  axis(2, at = seq( -0.2, 0.2, 0.05 ), lwd = 2, cex.lab = 1.2, las = 1, tck  = 0.02,
cex.axis = 1.5)
  axis(2, at = seq( -0.2, 0.2, 0.01 ), lwd = 2, cex.lab = 1.2, las = 1, tck  = 0.01,
cex.axis = 1.5, labels =FALSE)
  axis(4, at = seq(-0.2, 0.2, 0.05) , lwd = 2, labels = FALSE, tck  = 0.02)
  axis(4, at = seq( -0.2, 0.2, 0.01 ), lwd = 2, cex.lab = 1.2, las = 1, tck  = 0.01,
cex.axis = 1.5, labels =FALSE)
  title(ylab = expression(bold("Obs. - Theoret.,  ")~bolditalic(F(L/W)[obs]~ " -"
~F(L/W)[theo])),cex.lab = 1.5, line = 4)
  title(xlab = expression(bold("Length to Width Ratio, " ~bolditalic(L/W))), cex.lab = 1.5)
  points(max_x_d, all_d[maxx], col = "red", pch = 8)
  label = round(max_x_d, 2)
  textxy(max_x_d, all_d[maxx],(expression(~bolditalic("L/W ="))), cex = 1.5, col = "red")
  div = ((max(data)+5)/7)
  textxy((max_x_d +div), all_d[maxx],label, cex = 1.5, col = "red")
  legend("topright", do.call("expression", list(location,  phrase, phrase_a, "Cumulative
Frequency Difference", "Max D value")), lty = c(NA,NA,NA, NA, NA), col = c(NA,NA,  NA,
"darkgreen", "red"), pch = c(NA,NA, NA,20, 8), lwd = c(NA,NA, NA,NA, NA),bty = 'o', cex =
1.1, bg ="white", box.lwd = 2, inset = 0.02)

  # Use monte carlo modelling to test goodness of fit
  testresult = numeric(2500)
  dval = numeric(2500)

  rm(mle)
  rm(params)
  #Set up for loop
  for (i in 1:2500){
    # Generate n random values of LWR using the parameters fit to the observed data. N
should be the same as the number of observed LWRS.
    rand_lwrs = (rpearsonV(n, shape_rho,  location_s,  (1/scale_a)))
    rand_lwrs_cumulative = sort(rand_lwrs)

    # Use MLE to fit an inverse gamma to these randomly generated LWRs
    # Optimise the parameter values
    mle = optim(c(1.4,0.001,-0.0001),ll)
    params = mle$par
    shape_f = params [1]
    #Scale is the reciprocal of the actual scale value
    scale_f = 1/(params[2])
    location_f = params [3]
    parset = list(shape_f, location_f, scale_f)
    teo_cdf = ppearsonV(rand_lwrs_cumulative, params = parset, lower.tail = FALSE)
    rand_cdf = (1:n)/n
    rand_cdf = sort(rand_cdf, decreasing = TRUE)
    rand_D= max(abs(rand_cdf - teo_cdf))
    dval[i] = rand_D
    if (rand_D <= empirical_D){ testresult[i] = 0}
    if (rand_D > empirical_D){ testresult[i] = 1}}

  # The number of times the Empirical D value is smaller than the Monte carlo D value ###
  sum(testresult)
  # Proportion of times the D value is smaller than the Monte Carlo D value ###
  Proportion = sum(testresult) / 2500

  # Plot a boxplot of all D-values from Monte Carlo Techinque and plot empirical D value
  par(mar= c(5,6,2,2))
  yupper = (round(max(dval), 2)) + 0.1
  yupper2 = round(yupper, 1)
  boxplot(dval, range = 0, lwd = 2, lty = 1, axes = FALSE, xaxs ="i", yaxs = "i", ylim =
c(0, yupper2) )
  axis(1, tck = 0, labels = FALSE, at = c(0,2), lwd = 2)
  axis(2, lwd = 2, at = seq(0, yupper2, 0.05), las = 1, cex.axis = 1.5, tck = 0.02)
  axis(2, lwd = 2, at = seq(0, yupper2, 0.01), las = 1, cex.axis = 1.5, tck = 0.01, labels
=FALSE)
  axis(3, tck = 0, labels = FALSE, at = c(0,2), lwd = 2)
  axis(4, lwd = 2, at = seq(0, yupper2, 0.05), las = 1, cex.axis = 1.5, labels = FALSE, tck
= 0.02)
  axis(4, lwd = 2, at = seq(0, yupper2, 0.01), las = 1, cex.axis = 1.5, tck = 0.01, labels
=FALSE)
```

```
  title(ylab = expression(bold(D - Value)), cex.lab = 1.5, line = 4)
  abline((quantile(dval, 0.9)), 0, lty = 2, col = "grey", lwd = 2)
  abline((quantile(dval, 0.95)), 0, lty = 3, col = "lightgrey", lwd = 2)
  legend("topright", do.call("expression", list(location, phrase, phrase_a, "95th
Percentile", "90th Percentile","Empirical D Value")), lty = c(NA,NA, NA,3,2,1), col =
c(NA,NA, NA, "lightgrey", "grey", "red"), pch = c(NA, NA, NA, NA, NA, ">"), lwd = c(NA,NA,
NA, 2,2,2),bty = 'o', cex = 1.1, bg="white",box.lwd = 2, inset = 0.02)
  arrows((1-0.1), empirical_D, (1-0.01),empirical_D, col = "red", length = 0.15, lwd = 3)

  # Now produce bootstapped samples of the real data and fitting to each sample
  startshape = shape_rho
  startscale = scale_a
  startlocation = location_s
  estsa <- sapply(1:1000, function(i) {

  #Sample the data with replacement (the bootstrapping part)
    xi <- sample(data, size = n, replace = TRUE)
    ll = function(par){
    if(par[1]>0 & par[2]>0 & par[3]<min(xi)) return( -sum(log(dinvgamma(xi-
par[3],par[1],par[2]))) )
    else return(Inf)}
    optim(c(startshape, startscale, startlocation),ll)})

  bootstrap_shape = numeric(1000)
  bootstrap_location = numeric(1000)
  bootstrap_scale = numeric(1000)
  bootstrap_rollover = numeric(1000)
  for (b in 1:1000){
  bootstrap_shape [b] = estsa[,b]$par[1]
  bootstrap_location[b] = estsa[,b]$par[3]
  bootstrap_scale[b]  = estsa[,b]$par[2]}
  bootstrap_rollover[b] = (bootstrap_location[b]/(bootstrap_shape [b] + 1))+
bootstrap_scale[b]
  bootstrap_scale_recip = 1/bootstrap_scale

  m = matrix(c(bootstrap_shape, bootstrap_location, bootstrap_scale_recip), nrow= 1000, ncol
=3)
  ests = apply(m, 1, function(x) dpearsonV(xs, params = x))

  #Plot the first distritbution
  par(mar= c(5,6,2,2))
  plot(xs, ests[,1], type = "l", col="white", ylim= c(10^-8, 10), log = "xy", ylab = "",
xlab=expression(bold("Length to Width Ratio, "~bolditalic(L/W))), axes = FALSE, yaxs ="i",
xaxs = "i", cex.lab = 1.7 )
  ticks = c(0.1,1,2,3,4,5,6,7,8,9,10,20,30,40,50,60,70,80,90,100)
  title(ylab = expression(bold("Probability Density, " ~bolditalic(p(L/W)))), cex.lab = 1.7,
line = 4)
  axis(1, at = c(1,10, 100), labels =( expression(~bold(10^0),~bold(10^1), ~bold(10^2))),
lty = 1, lwd = 0, tck = 0, cex.axis = 1.7)
  yticks = c(0.000000001, 0.000000002, 0.000000003, 0.000000004, 0.000000005,  0.000000006,
0.000000007, 0.000000008, 0.000000009, 0.00000001, 0.00000002, 0.00000003, 0.00000004,
0.00000005,  0.00000006, 0.00000007, 0.00000008, 0.00000009,0.0000001, 0.0000002, 0.0000003,
0.0000004, 0.0000005,  0.0000006, 0.0000007, 0.0000008, 0.0000009, 0.000001, 0.000002,
0.000003,  0.000004, 0.000005, 0.000006, 0.000007, 0.000008, 0.000009, 0.00001, 0.00002,
0.00003, 0.00004, 0.00005, 0.00006, 0.00007, 0.00008, 0.00009,0.0001, 0.0002, 0.0003,
0.0004, 0.0005, 0.0006, 0.0007, 0.0008, 0.0009, 0.001, 0.002, 0.003, 0.004, 0.005, 0.006,
0.007, 0.008, 0.009, 0.01, 0.02, 0.03, 0.04, 0.05, 0.06, 0.07, 0.08, 0.09, 0.1, 0.2, 0.3,
0.4,0.5,0.6,0.7,0.8,0.9,1,2,3,4,5,6,7,8,9,10)
  axis(2, at = c(0.000000001, 0.00000001,0.0000001, 0.000001, 0.00001, 0.0001, 0.001, 0.01,
0.1, 1,10), labels =(expression(~bold(10^-9),~bold(10^-8),~bold(10^-7),~bold(10^-
6),~bold(10^-5), ~bold(10^-4), ~bold(10^-3), ~bold(10^-2), ~bold(10^-1), ~bold(10^0) ,
~bold(10^1))), lty = 1, lwd = 0, tck = 0, cex.axis = 1.7, las = 1)

  # Plot add cdfs calcuated from each bootstrap
  for(i in 2:500)
  lines(xs, ests[, i], col=rgb(.6, .6, .6, .1))
  quants <- apply(ests, 1, quantile, c(0.025, 0.5, 0.975))
  median_pdf = quants[2,]
  lower_pdf = quants[3,]
  upper_pdf = quants[1,]
  lines(xs, quants[1, ], col="red", lwd=1.5, lty=2)
```

```
  lines(xs, quants[3, ], col="red", lwd=1.5, lty=2)
  lines(xs, quants[2, ], col="darkred", lwd=2)
  axis(1, ticks, labels = NA, lty = 1, lwd = 2, tck = 0.01)
  axis(2, yticks, labels = NA, lty = 1, lwd = 2, tck = 0.01)
  axis(3, ticks, labels = NA, lty = 1, lwd = 2, tck = 0.01)
  axis(4, yticks, labels = NA, lty = 1, lwd = 2, tck = 0.01)
  legend("topright", do.call("expression", list(location, phrase, phrase_a, "Fit to
bootstrapped samples", "Median fit", "5th/95th Percentile Fit")), lty = c(NA,NA,NA, 1,1, 2),
col = c(NA,NA,NA, "Grey", "red", "red"),  lwd = c(NA,NA,NA, 0.5,2,1),bty = 'o', cex = 1.1,
box.lwd = 2, inset = 0.01)
  points(emp_bin_mids, emp_prob_dens, pch = 20)

  bootstrap_rollover = numeric(1000)
  for(i in 1:1000){
    max_pdf = max(ests[,i])
    loc_max = match(max_pdf, ests[,i])
    location_1 = xs[loc_max]
    bootstrap_rollover[i] = location_1}

  # Save data from this iteration and start next loop
  variable1 = paste(run, "shape", location, sep = "")
  assign(variable1, bootstrap_shape)
  variable2 = paste(run,"location", location, sep = "")
  assign(variable2, bootstrap_location)
  variable3 = paste(run,"scale", location, sep = "")
  assign(variable3, bootstrap_scale)
  variable4 = paste(run,"rollover", location, sep = "")
  assign(variable4, bootstrap_rollover)
  variable5 = paste(run,"emp_d", location, sep = "")
  assign(variable5, empirical_D)
  variable10 = paste(run,"d_val", location, sep = "")
  assign(variable10, dval)
  variable6 = paste(run,"dprop", location, sep = "")
  assign(variable6, Proportion)
  variable7 = paste(run,"median_pdf", location, sep = "")
  assign(variable7, median_pdf)
  variable8 = paste(run,"upper_pdf", location, sep = "")
  assign(variable8, upper_pdf)
  variable9 = paste(run,"lower_pdf", location, sep = "")
  assign(variable9, lower_pdf)
  variable11 = paste(run,"emp_pdf", location, sep = "")
  assign(variable11, emp_prob_dens)
  variable12 = paste(run,"emp_mids", location, sep = "")
  assign(variable12, emp_bin_mids)
  # Start new loop to analyse next landlside area category
}}

#### 3. Produce summary plots for all landslide area categories
# Define a colour for each landslide area category
LWRrainbow_pal = c("#902525", "#ff0707", "#f18c37", "#fcfc02", "#91cd4e", "#1EBB66",
"#07b1f0", "#0069bd", "#001055", "#68259b")

# Create a boxplot of L/W within each area category. First create a blank graph with a dummy
data series
par( family = "serif", mfrow=c(1,1),mar= c(9,6,1,1), oma = c(1,0,0,0),  lwd = 2, tck = 0.03)
synth = c(0.01,0.05,0.01,0.02)
boxplot(synth, range = 0, border = "white", col = "white", lwd = 1, lty = 1, axes = FALSE,
xaxs ="i", yaxs = "i", ylim = c(0, 0.5), xlim = c(0,11))
axis(1, tck = 0.03, labels = FALSE, at = c(seq(-1,12)), lwd = 2)
axis(2, lwd = 2, at = seq(0, 0.5, 0.05), las = 1, cex.axis = 1.5, tck = 0.03)
axis(3, tck = 0.03, labels = FALSE, at = c(seq(-1,12)), lwd = 2)
axis(4, lwd = 2, at = seq(0, 0.5, 0.05), las = 1, cex.axis = 1.5, labels = FALSE, tck =
0.03)
title(ylab = expression(bold(D - Value)), cex.lab = 1.5, line = 4)
labels = list( phrase1, phrase2, phrase3, phrase4, phrase5, phrase6, phrase7, phrase8,
phrase9, phrase10)
text(1:10, par("usr")[3]-0.01, srt = 45, adj = 1,labels = do.call("expression", labels), xpd
= TRUE)
title(xlab = expression(bold("Landslide Area Category, ") ~bolditalic(C[A[L] ])
~bold((m^2))), cex.lab = 1.5, line = 8)
```

```
# Add boxplots of L/W one by one for each landslide area category
for (i in 1:10){
  data = alldata[i]
  data = data[!is.na(data)]
  if ((length(data))<1){
    next
  } else {
  name = paste("LW",i,"d_val",location, sep = "")
  box = get(name)
  boxplot(box, range = 0, lwd = 2, lty = 1, border =LWRrainbow_pal[i],  add = TRUE, at = i ,
axes = FALSE)
  points((i), (quantile(box, 0.9)), pch = "-", col = "black", lwd = 1, cex = 3)
  points((i), (quantile(box, 0.95)), pch = "-", col = "gray", lwd = 1, cex = 3)
  name = paste("LW", i,"emp_d", location, sep = "")
  arrow =  get(name)
  arrows((i-0.3), arrow, (i-0.25),arrow, col = "black", length = 0.15)
}}
legend("topright", do.call("expression", list(location, "95th Percentile", "90th
Percentile","Empirical D Value")), lty = c(NA,1,1,1), col = c(NA, "black", "grey", "black"),
pch = c(NA, NA, NA, ">"), lwd = c( NA, 3,3,NA),bty = 'o',, box.col = "black", bg =
rgb(1,1,1,0.5), cex = 1.1, box.lwd = 2, inset = 0.01)

# Create plot of inverse gamma probability density function fit to each landslide area
category. First create a blank graph with a dummy data series.
par(mfrow = c(1,1),mar = c(4,6,1,2), family = "serif")
linetype = c(1,1,1,5,5,5,5,3,3,3)
plot(1,1, type = "l", col = "white", lwd = 3, xlim = c(1,20), ylim= c(10^-6, 10), log =
"xy", ylab = "", xlab = "",  axes = FALSE, yaxs ="i", xaxs = "i", cex.lab = 1.7 )
ticks = c(0.1,1,2,3,4,5,6,7,8,9,10,20,30,40,50,60,70,80,90,100)
title(ylab = expression(bold("Probability Density, " ~bolditalic(p(L/W)))), cex.lab = 1.7,
line = 3.2)
title(xlab=expression(bold("Length to Width Ratio, "~bolditalic(L/W))), cex.lab = 1.7, line
= 3)
axis(1, at = c(1,10, 19), labels =( expression(~bold(10^0),~bold(10^1), ~bold(2 ~"×"
~10^1))), lty = 1, lwd = 0, tck = 0, cex.axis = 1.6)
yticks = c(0.000000001, 0.000000002, 0.000000003, 0.000000004, 0.000000005,  0.000000006,
0.000000007, 0.000000008, 0.000000009, 0.00000001, 0.00000002, 0.00000003, 0.00000004,
0.00000005,  0.00000006, 0.00000007, 0.00000008, 0.00000009,0.0000001, 0.0000002, 0.0000003,
0.0000004, 0.0000005,  0.0000006, 0.0000007, 0.0000008, 0.0000009, 0.000001, 0.000002,
0.000003,  0.000004, 0.000005, 0.000006, 0.000007, 0.000008, 0.000009, 0.00001, 0.00002,
0.00003, 0.00004, 0.00005, 0.00006, 0.00007, 0.00008, 0.00009,0.0001, 0.0002, 0.0003,
0.0004, 0.0005, 0.0006, 0.0007, 0.0008, 0.0009, 0.001, 0.002, 0.003, 0.004, 0.005, 0.006,
0.007, 0.008, 0.009, 0.01, 0.02, 0.03, 0.04, 0.05, 0.06, 0.07, 0.08, 0.09, 0.1, 0.2, 0.3,
0.4,0.5,0.6,0.7,0.8,0.9,1,2,3,4,5,6,7,8,9,10)
axis(2, at = c(0.000000001, 0.00000001,0.0000001, 0.000001, 0.00001, 0.0001, 0.001, 0.01,
0.1, 1,10), labels =(expression(~bold(10^-9),~bold(10^-8),~bold(10^-7),~bold(10^-
6),~bold(10^-5), ~bold(10^-4), ~bold(10^-3), ~bold(10^-2), ~bold(10^-1), ~bold(10^0) ,
~bold(10^1))), lty = 1, lwd = 0, tck = 0, cex.axis = 1.7, las = 1)
axis(1, ticks, labels = NA, lty = 1, lwd = 2, tck =0.015)
axis(2, yticks, labels = NA, lty = 1, lwd = 2, tck = 0.015)
axis(3, ticks, labels = NA, lty = 1, lwd = 2, tck = 0.015)
axis(4, yticks, labels = NA, lty = 1, lwd = 2, tck = 0.015)
LWRrainbow_pal_transp = c("#90252510", "#ff070710", "#f18c3710", "#fcfc0210", "#91cd4e10",
"#1EBB6610", "#07b1f010", "#0069bd10", "#00105510", "#68259b10")

# Plot pdfs one by one for each landslide area category
for (i in 10:1){
  data = alldata[i]
  data = data[!is.na(data)]
  if ((length(data))<2){
    next
  } else {
    uppol_name = paste("LW",i,"upper_pdf",location, sep = "")
    uppol_name = get(uppol_name)
    lopol_name = paste("LW",i,"lower_pdf",location, sep = "")
    lopol_name = get(lopol_name)
    polygon(c(xs, rev(xs)), c(uppol_name, rev(lopol_name)) , col = LWRrainbow_pal_transp[i],
     border = LWRrainbow_pal_transp[i])
    med_pdf_name = paste("LW",i,"median_pdf",location, sep = "")
    med_pdf_name = get(med_pdf_name)
    points(xs, med_pdf_name, type = "l", lty = linetype[i], lwd = 2, col=LWRrainbow_pal[i])
```

```
  }}
val <- substitute("value" == phrase, list(phrase = phrase))
legend("topright", do.call("expression", list(location, phrase1, phrase2, phrase3, phrase4,
phrase5, phrase6, phrase7, phrase8, phrase9, phrase10)), lty = c(NA,linetype), col = c(NA,
LWRrainbow_pal),  lwd = (3),bty = 'o', cex = 1, box.lwd = 2, inset = 0.03, box.col =
rgb(0,0,0))

# Create individual plots for pdf and observed probabiltiy densities for each landslide area
category
par(oma = c(5,5,1,1), mfrow=c(5,2), family = "serif", mar= c(2,3.5,1,1), lwd = 2, tck =
0.03)
for (i in 1:10){
  data = alldata[i]
  data = data[!is.na(data)]
  if ((length(data))<1){
        plot(1,1, type = "l", col = "white", lwd = 3, xlim = c(1,20), ylim= c(10^-6, 10),
        log = "xy", ylab = "", xlab = "",  axes = FALSE, yaxs ="i", xaxs = "i", cex.lab =
        1.7, bg = "lightgrey" )
    ticks = c(0.1,1,2,3,4,5,6,7,8,9,10,20,30,40,50,60,70,80,90,100)
    axis(1, at = c(1,10, 19), labels =( expression(~bold(10^0),~bold(10^1), ~bold(2 ~"x"
    ~10^1))), lty = 1, lwd = 0, tck = 0, cex.axis = 1.4)
    yticks = c(0.000000001, 0.000000002, 0.000000003, 0.000000004, 0.000000005,
    0.000000006, 0.000000007, 0.000000008, 0.000000009, 0.00000001, 0.00000002, 0.00000003,
    0.00000004, 0.00000005,  0.00000006, 0.00000007, 0.00000008, 0.00000009,0.0000001,
    0.0000002, 0.0000003, 0.0000004, 0.0000005,  0.0000006, 0.0000007, 0.0000008,
    0.0000009, 0.000001, 0.000002, 0.000003,  0.000004, 0.000005, 0.000006, 0.000007,
    0.000008, 0.000009, 0.00001, 0.00002, 0.00003, 0.00004, 0.00005, 0.00006, 0.00007,
    0.00008, 0.00009,0.0001, 0.0002, 0.0003, 0.0004, 0.0005, 0.0006, 0.0007, 0.0008,
    0.0009, 0.001, 0.002, 0.003, 0.004, 0.005, 0.006, 0.007, 0.008, 0.009, 0.01, 0.02,
    0.03, 0.04, 0.05, 0.06, 0.07, 0.08, 0.09, 0.1, 0.2, 0.3,
    0.4,0.5,0.6,0.7,0.8,0.9,1,2,3,4,5,6,7,8,9,10)
    axis(2, at = c(0.000000001, 0.00000001,0.0000001, 0.000001, 0.00001, 0.0001, 0.001,
    0.01, 0.1, 1,10), labels =(expression(~bold(10^-9),~bold(10^-8),~bold(10^-7),~bold(10^-
    6),~bold(10^-5), ~bold(10^-4), ~bold(10^-3), ~bold(10^-2), ~bold(10^-1), ~bold(10^0) ,
    ~bold(10^1))), lty = 1, lwd = 0, tck = 0, cex.axis = 1.4, las = 1)
    axis(1, ticks, labels = NA, lty = 1)
    axis(2, yticks, labels = NA, lty = 1)
    axis(3, ticks, labels = NA, lty = 1)
    axis(4, yticks, labels = NA, lty = 1)
    text(5, 0.01, "(Null)", col = LWRrainbow_pal[i], cex = 2)
    catname = paste("phrase",i, sep = "")
    catname = get(catname)
     legend("topright", do.call("expression", list(catname)), lty = c(NA), col = c(NA), pch
     = c(NA), lwd = c(NA),bty = 'o', cex = 1, box.lwd = 0.5, box.col = "lightgrey", bg =
     rgb(1,1,1,0.5),  inset = 0.01)
    next
  } else {
    rollover_name = paste("LW",i,"rollover",location, sep = "")
    rollover_name = get(rollover_name)
    med_rollover = median(rollover_name)
    lower_rollover = quantile(rollover_name, 0.05)
    upper_rollover = quantile(rollover_name, 0.95)
    print(LWRrainbow_pal[i])
    uppol_name = paste("LW",i,"upper_pdf",location, sep = "")
    uppol_name = get(uppol_name)
    lopol_name = paste("LW",i,"lower_pdf",location, sep = "")
    lopol_name = get(lopol_name)
     plot(1,1, type = "l", col = "white", lwd = 3, xlim = c(1,20), ylim= c(10^-6, 10), log =
     "xy", ylab = "", xlab = "",  axes = FALSE, yaxs ="i", xaxs = "i", cex.lab = 1.7 )
    print(LWRrainbow_pal_transp[i])
     polygon(c(xs, rev(xs)), c(uppol_name, rev(lopol_name)) , col =
     LWRrainbow_pal_transp[i],border = LWRrainbow_pal_transp[i])
    abline(v =  med_rollover, col = LWRrainbow_pal[i], lwd = 2, lty = 2)
    print(LWRrainbow_pal_transp[i])
     rect(lower_rollover, 0.000001, upper_rollover, 10, col = LWRrainbow_pal_transp[i],
     border = LWRrainbow_pal[i], lty = 3 )
    med_pdf_name = paste("LW",i,"median_pdf",location, sep = "")
    med_pdf_name = get(med_pdf_name)
    points(xs, med_pdf_name, type = "l", lty = 1, lwd = 2, col=LWRrainbow_pal[i])
    empy_name = paste("LW",i,"emp_pdf",location, sep = "")
    empy_name= get(empy_name)
```

```
    empx_name = paste("LW",i,"emp_mids",location, sep = "")
    empx_name = get(empx_name)
    print(i)
    points(empx_name, empy_name, type = "p", col = LWRrainbow_pal[i], pch = 20, cex = 3)
    ticks = c(0.1,1,2,3,4,5,6,7,8,9,10,20,30,40,50,60,70,80,90,100)
    axis(1, at = c(1,10, 19), labels =( expression(~bold(10^0),~bold(10^1), ~bold(2 ~"×"
    ~10^1))), lty = 1, lwd = 0, tck = 0, cex.axis = 1.4)
    yticks = c(0.000000001, 0.000000002, 0.000000003, 0.000000004, 0.000000005,
    0.000000006, 0.000000007, 0.000000008, 0.000000009, 0.00000001, 0.00000002, 0.00000003,
    0.00000004, 0.00000005,  0.00000006, 0.00000007, 0.00000008, 0.00000009,0.0000001,
    0.0000002, 0.0000003, 0.0000004, 0.0000005,  0.0000006, 0.0000007, 0.0000008,
    0.0000009, 0.000001, 0.000002, 0.000003,  0.000004, 0.000005, 0.000006, 0.000007,
    0.000008, 0.000009, 0.00001, 0.00002, 0.00003, 0.00004, 0.00005, 0.00006, 0.00007,
    0.00008, 0.00009,0.0001, 0.0002, 0.0003, 0.0004, 0.0005, 0.0006, 0.0007, 0.0008,
    0.0009, 0.001, 0.002, 0.003, 0.004, 0.005, 0.006, 0.007, 0.008, 0.009, 0.01, 0.02,
    0.03, 0.04, 0.05, 0.06, 0.07, 0.08, 0.09, 0.1, 0.2, 0.3,
    0.4,0.5,0.6,0.7,0.8,0.9,1,2,3,4,5,6,7,8,9,10)
    axis(2, at = c(0.000000001, 0.00000001,0.0000001, 0.000001, 0.00001, 0.0001, 0.001,
    0.01, 0.1, 1,10), labels =(expression(~bold(10^-9),~bold(10^-8),~bold(10^-7),~bold(10^-
    6),~bold(10^-5), ~bold(10^-4), ~bold(10^-3), ~bold(10^-2), ~bold(10^-1), ~bold(10^0) ,
    ~bold(10^1))), lty = 1, lwd = 0, tck = 0, cex.axis = 1.4, las = 1)
    axis(1, ticks, labels = NA, lty = 1)
    axis(2, yticks, labels = NA, lty = 1)
    axis(3, ticks, labels = NA, lty = 1)
    axis(4, yticks, labels = NA, lty = 1)
    n = length(data)
    faith = bquote(bolditalic("n = " ~bold(.(n))))
    catname = paste("phrase",i, sep = "")
    catname = get(catname)
    val <- substitute("value" == phrase, list(phrase = phrase))
    legend("topright", do.call("expression", list(catname, faith)), lty = c(NA, NA), col =
    c(NA, NA), pch = c(NA,  NA), lwd = c(NA, NA),bty = 'o', cex = 1, box.lwd = 0.5, box.col
    = "lightgrey", bg = rgb(1,1,1,0.5),  inset = 0.01)
}}
title(ylab = expression(bold("Probability Density, " ~bolditalic(p(L/W)))), cex.lab = 3,
line = 0, outer  = TRUE)
title(xlab=expression(bold("Length to Width Ratio, "~bolditalic(L/W))), cex.lab = 3, line =
2.2, outer  = TRUE)

# Perform a KS test to compare distribution of L/W within each landslide area category
par(mar = c(14,14,1,1), family = "serif")

#Set up empty matrices to store data
d_table = matrix(,nrow = 10, ncol =10)
p_table = matrix(,nrow = 10, ncol =10)
rownames(p_table) = c("LW1", "LW2", "LW3", "LW4", "LW5", "LW6", "LW7", "LW8", "LW9", "LW10")
colnames(p_table) = c("LW1", "LW2", "LW3", "LW4", "LW5", "LW6", "LW7", "LW8", "LW9", "LW10")
rownames(d_table) = c("LW1", "LW2", "LW3", "LW4", "LW5", "LW6", "LW7", "LW8", "LW9", "LW10")
colnames(d_table) = c("LW1", "LW2", "LW3", "LW4", "LW5", "LW6", "LW7", "LW8", "LW9", "LW10")

# Extract data for each landslide area category
for (A in 1:10){
  data = alldata[A]
  data = data[!is.na(data)]

  if ((length(data))<1){
    next
  } else {
  foo = paste("outer = ", paste(A))

  # Perform KS test between each landslide area category
  for (j in (A+1):10){
    compare = alldata[j]
    compare = compare[!is.na(compare)]
    if ((length(compare))<1){
      next
    } else {
      foo = paste("inner = ", paste(j))
      kstest = ks.test(data, compare)
      d_table[A,j] = kstest$statistic
      p_table[A,j] = kstest$p
```

```
}}}}

p_value_matrix = matrix(0,nrow = 10, ncol =10)
for (A in 1:10){
  data = alldata[A]
  data = data[!is.na(data)]

  if ((length(data))<1){
    p_value_matrix [,A] = 1
    p_value_matrix [A,] = 1
  }}
for (x in 1:9){
  for(y in (x+1):10){
    p_value_matrix [x,y] = 2
}}

pal2 = colorRampPalette(c("white", "lightgrey"))(n =3)
my_palette <- colorRampPalette(c( "#91CF60", "#FFFFBF", "#FC8D59"))(5)
col_breaks <- c(0,0.01,0.05,0.1,0.25,1)

rounded_p_table = round(p_table,2)
rounded_p_table[10,10]=NA

# Produce plot of KS test p_values where colour corresponds to p-value
image(1:10, 1:10, p_value_matrix, col = pal2, axes = FALSE, xlab = "", ylab = "")
image(1:ncol(rounded_p_table),1:nrow(rounded_p_table),t(rounded_p_table), col = my_palette,
breaks = col_breaks, axes = FALSE, xlab = "", ylab = "", add = TRUE)
axis(1, at = seq(-1,11,1), lwd = 2, tck = 0.0, labels =FALSE)
axis(2, at = seq(-1,11,1), lwd = 2, tck = 0.0, labels =FALSE)
axis(3, at = seq(-1,11,1), lwd = 2, tck = 0.0, labels =FALSE)
axis(4, at = seq(-1,11,1), lwd = 2, tck = 0.0, labels =FALSE)
axis(1, at = seq(-1.5,11.5,1), lwd= 2, tck = 0.03, labels = FALSE)
axis(2, at = seq(-1.5,11.5,1), lwd= 2, tck = 0.03, labels = FALSE)
axis(3, at = seq(-1.5,11.5,1), lwd= 2, tck = 0.03, labels = FALSE)
axis(4, at = seq(-1.5,11.5,1), lwd= 2, tck = 0.03, labels = FALSE)
for (x in 1:ncol(rounded_p_table))
  for (y in 1:nrow(rounded_p_table))
    text(x, y, rounded_p_table[y,x])
labels = list( phrase1, phrase2, phrase3, phrase4, phrase5, phrase6, phrase7, phrase8,
phrase9, phrase10)
text(1:10, par("usr")[3] - 0.25, srt = 45, adj = 1,labels = do.call("expression", labels),
xpd = TRUE)

dev.off()
```

Appendix B: Landslide-Road Impact Model (LRIM) Python Model Code

Below is the Python code which executes in Grass GIS 7.0 for the Landslide-Road Impact Model (LRIM). It has an optional element of randomly selecting portions of the road network to buffer and adjust the susceptibility within those zones.

```
#!/usr/bin/env python
############################################################################################
#####
#                                                                                          #
# MODULE:    v.lrim                                                                        #
#                                                                                          #
# AUTHOR(S):   F. Taylor, B. Malamud, M. Santangelo, I. Marchesini, F. Guzzetti            #
# PURPOSE:     Simulate triggered landslide event inventory impact on road network         #
#                                                                                          #
# DATE:        1st April 2015                                                              #
# COPYRIGHT:   (C) 2015 F. Taylor, B. Malamud, M. Santangelo, I. Marchesini, F. Guzzetti   #
#                                                                                          #
#       This program is free software under the GNU General Public                         #
#       License (>=v2). Read the file COPYING that comes with GRASS                        #
#       for details.                                                                       #
#                                                                                          #
############################################################################################
#####

# User specifies input files
#%module
#% description: randomly simulate landslides and hit roads. The outputs vectors are called
intersetion1 and intersection2. Inputs must be in UTM projection.
#% keywords: vector
#% keywords: landslide
#%end
#%option
#% key: roads
#% type: string
#% gisprompt: old,vector,vector
#% description: name of the road network layer
#% required: yes
#%end
#%option
#% key: buffered_roads50
#% type: string
#% gisprompt: old,vector,vector
#% description: name of the road network layer with 50 m buffer
#% required: yes
#%end
#%option
#% key: buffered_roads100
#% type: string
#% gisprompt: old,vector,vector
#% description: name of the road network layer with 100 m buffer
#% required: yes
#%end
#%option
#% key: density
#% type: double
#% description: density of landslides as number per square kilometer
#% required: yes
#% answer:
#%end
```

```
#%option
#% key: study_area
#% type: string
#% gisprompt: old,vector,vector
#% description: study area
#% required: yes
#%end
#%option
#% key: susceptibility_map_vect
#% type: string
#% gisprompt: old,vector,vector
#% description: Vector shapefile of susceptibility
#% required: yes
#%end
#%option
#% key: dem10
#% type: string
#% gisprompt: old,raster,dcell
#% description: 10m2 resolution digital elevation model
#% required: yes
#%end
#%option
#% key: aspect_map
#% type: string
#% gisprompt: old,raster,dcell
#% description: aspect map
#% required: yes
#%end
#%option
#% key: num_iterations
#% type: double
#% description: number of iterations to run the model for
#% required: yes
#% answer: 100
#%end

# Optional additional road maps by hierarchy, used for random buffering of roads to
increase/decrease # susceptibility
#% key: motorway_roads
#% type: string
#% gisprompt: old,vector,vector
#% description: name of the trunk road network layer
#% required: yes
#%end
#%option
#% key: secondary_roads
#% type: string
#% gisprompt: old,vector,vector
#% description: name of the secondary road network layer
#% required: yes
#%end
#%option
#% key: local_roads
#% type: string
#% gisprompt: old,vector,vector
#% description: name of the tertiary road network layer
#% required: yes
#%end
#%option
#% key: minor_roads
#% type: string
#% gisprompt: old,vector,vector
#% description: name of the minor road network layer
#% required: yes
#%end

# Import necessary modules
from __future__ import division
import numpy
import random
import scipy
```

```
from scipy.stats import invgamma
from numpy import *
import shlex
import sys
import os
import atexit
import grass.script as grass
from grass.script import core as grasscore

def cleanup():
     pass

def main():
     # Assign each input file as a variable
     roads = options['roads']
     slide_density1 = options['density']
     slide_density = float(slide_density1)
     study_area = options['study_area']
     susceptibility_map_vect = options['susceptibility_map_vect']
     dem10 = options['dem10']
     num_iterations1 = options['num_iterations']
     num_iterations2 = float(num_iterations1)
     num_iterations = int(num_iterations2)
     aspect_map = options['aspect_map']
     buffered_roads50 = options['buffered_roads50']
     buffered_roads100 = options['buffered_roads100']

     motorway_roads = options['motorway_roads']
     secondary_roads = options['secondary_roads']
     local_roads = options['local_roads']
     minor_roads = options['minor_roads']

     #Remove temporary files
     grass.run_command("g.mremove",vect="point*",flags = "f")
     grass.run_command("g.mremove",vect="road_blocks*",flags = "f")
     grass.run_command("g.mremove",vect="random_points*",flags = "f")
     grass.run_command("g.mremove",vect="ind_landslide*",flags = "f")
     grass.run_command("g.mremove",vect="landslide*",flags = "f")

     grass.run_command("g.mremove",vect="final_points*",flags = "f")
     grass.run_command("g.mremove",vect="set_points_drop*", = "f")
     grass.run_command("g.mremove",vect="points_in_area*",flags = "f")
     grass.run_command("g.mremove",vect="points_in_area*",flags = "f")

     grass.run_command("g.mremove",vect = "polygons_neighbors*",flags = "f")
     grass.run_command("g.mremove",vect = "landslide_inventory_uncleaned*",flags = "f")
     grass.run_command("g.mremove",vect = "set_points_drop*",flags = "f")

     grass.run_command("g.mremove",vect = "landslide_inventory_patch*",flags = "f")
     grass.run_command("g.remove", vect = "motorway_roads_rand_selection")
     grass.run_command("g.remove", vect = "motorway_roads_rand_selection_buffered")
     grass.run_command("g.remove", vect = "secondary_roads_rand_selection")
     grass.run_command("g.remove", vect = "secondary_roads_rand_selection_buffered")
     grass.run_command("g.remove", vect = "local_roads_rand_selection")
     grass.run_command("g.remove", vect = "local_roads_rand_selection_buffered")
     grass.run_command("g.remove", vect = "minor_roads_rand_selection")
     grass.run_command("g.remove", vect = "minor_roads_rand_selection_buffered")
     grass.run_command("g.remove", vect = "patched_buffered_roads_all")
     grass.run_command("g.remove", vect = "overlay_susc_rds")
     grass.run_command("g.remove", vect = "buffered_roads_all_studyreg")
     grass.run_command("g.remove", rast = "buffered_roads_susc_multiplier")

     #Set region from view extent and set the resolution to 1 metre squared
     grass.run_command("g.region",
                       vect = study_area,
                       res = 1)

     #Calculate an aspect map from the DEM
     grass.run_command("r.slope.aspect",
                            elevation= dem10,
                            aspect= "aspect_map")
```

```
#Add the average azimuth (dip direction) to the susceptibility map
grass.run_command("v.rast.stats",
                          vector= susceptibility_map_vect,
                          raster= "aspect_map",
                          column_prefix="asp",
                          flags = "c")

# Set up a buffered susceptibilty map where there is a 500 m border around the map
grass.run_command("v.buffer",
                         input=study_area,
                         output= "buf_stud_reg",
                         distance=500,
                         minordistance=500)

grass.run_command("v.overlay",
                         ainput= "buf_stud_reg",
                         binput=susceptibility_map_vect,
                         operator= "or",
                         overwrite = True,
                         output= "overlay_susc")

grass.run_command("v.db.update",
                         map= "overlay_susc",
                         column="b_prob_to85c",
                         value=-1,
                         where="b_prob_to85c is NULL")

grass.run_command("v.db.update",
                         map= "overlay_susc",
                         column="b_aspect",
                         value=0,
                         where="b_aspect is NULL")

grass.run_command("v.db.dropcolumn",
                         map="overlay_susc",
                         columns="a_cat")

table_names = grass.read_command("v.db.connect",
                         flags = "c",
                         map="overlay_susc")

table_names = table_names.split('\n')
old_col_names = []
old_col_names.append(table_names[0])
new_col_names = []
new_col_names.append(table_names[0])
for k in range(1, len(table_names)):
     q = table_names[k].strip("INTEGER|")
     h = q.strip("DOUBLE PRECISION|")
     I = h.strip("CHARACTER|")
     J = I.replace("b_", "", 1)
     new_col_names.append(J)
     old_col_names.append(I)

for s in range(1, len(new_col_names)):
               colname = old_col_names[s]
               newcolname = new_col_names[s]
               grass.run_command("v.db.renamecolumn",
                         map="overlay_susc",
                         column= (colname,newcolname))

# Create a raster map of the susceptibility map
grass.run_command("v.to.rast",
                          input =  "overlay_susc",
                          output = "susc_rast",
                          use = "attr",
                          attrcolumn = "prob_to85c",
                          overwrite = True)

grass.run_command("v.to.rast",
```

```
                                input =  "overlay_susc",
                                 output = "susc_rast_cat",
                                 use = "attr",
                                 attrcolumn = "cat",
                                 overwrite = True)

    grass.run_command("v.to.rast",
                                 input =  "overlay_susc",
                                 output = "susc_rast_asp",
                                 use = "attr",
                                 attrcolumn = "aspect",
                                 overwrite = True)

    #Optional: Split the road network into 50 m segments in order to randomly select and buffer
later
    grass.run_command("v.split",
              input = motorway_roads,
              output = "motorway_roads_split",
              length = 50,
              units = "meters")

    grass.run_command("v.split",
              input = secondary_roads,
              output = "secondary_roads_split",
              length = 50,
              units = "meters")

    grass.run_command("v.split",
              input = local_roads,
              output = "local_roads_split",
              length = 50,
              units = "meters")

    grass.run_command("v.split",
              input = minor_roads,
              output = "minor_roads_split",
              length = 50,
              units = "meters")

    #v.split does not add in new categories for each road segment so file is exported and then
    imported again
    grass.run_command("v.out.ogr",
              input = "motorway_roads_split",
              dsn = "motorway_roads_split_out",
              format = "ESRI_Shapefile")

    grass.run_command("v.in.ogr",
              dsn  = "motorway_roads_split_out",
              output = "motorway_roads_split_cats")

    grass.run_command("v.out.ogr",
              input = "secondary_roads_split",
              dsn = "secondary_roads_split_out",
              format = "ESRI_Shapefile")

    grass.run_command("v.in.ogr",
              dsn  = "secondary_roads_split_out",
              output = "secondary_roads_split_cats")

    grass.run_command("v.out.ogr",
              input = "local_roads_split",
              dsn = "local_roads_split_out",
              format = "ESRI_Shapefile")

    grass.run_command("v.in.ogr",
              dsn  = "local_roads_split_out",
              output = "local_roads_split_cats")

    grass.run_command("v.out.ogr",
              input = "minor_roads_split",
              dsn = "minor_roads_split_out",
```

```
          format = "ESRI_Shapefile")

grass.run_command("v.in.ogr",
          dsn  = "minor_roads_split_out",
          output = "minor_roads_split_cats")


# The percentage of road length that should be buffered is specified
motorway_percentage = 88
secondary_percentage = 80
local_percentage = 81
minor_percentage = 91

motorway_length = grass.read_command("v.to.db",
     map = motorway_roads,
     option = "length",
     flags = 'c')

motorway_length = motorway_length.strip("total length|")
motorway_length = float(motorway_length)
motorway_length_to_be_buffered = (motorway_length/100) * motorway_percentage
motorway_num_sections_to_be_buffered = int(motorway_length_to_be_buffered / 50)

secondary_length = grass.read_command("v.to.db",
     map = secondary_roads,
     option = "length",
     flags = 'c')

secondary_length = secondary_length.strip("total length|")
secondary_length = float(secondary_length)
secondary_length_to_be_buffered = (secondary_length/100) * secondary_percentage
secondary_num_sections_to_be_buffered = int(secondary_length_to_be_buffered / 50)


local_length = grass.read_command("v.to.db",
     map = local_roads,
     option = "length",
     flags = 'c')

local_length = local_length.strip("total length|")
local_length = float(local_length)
local_length_to_be_buffered = (local_length/100) * local_percentage
local_num_sections_to_be_buffered = int(local_length_to_be_buffered / 50)


minor_length = grass.read_command("v.to.db",
     map = minor_roads,
     option = "length",
     flags = 'c')

minor_length = minor_length.strip("total length|")
minor_length = float(minor_length)
minor_length_to_be_buffered = (minor_length/100) * minor_percentage
minor_num_sections_to_be_buffered = int(minor_length_to_be_buffered / 50)

#Specify the width in meters to buffer each road type by
motorway_bufwidth = 22
secondary_bufwidth = 62
local_bufwidth = 41
minor_bufwidth = 51

#Take region values of RECTANGLE (not study area)
region = grasscore.region()
E = region['e']
W = region['w']
N = region['n']
S = region['s']

#Calculate the size of the study area from UTM coordinates
Study_area_width = E-W
Study_area_height = N-S
```

```
    # Add column to study area vector to add in calculated area in km2
    grass.run_command("v.db.addcolumn",
                      map = study_area,
                      col = "area_sqkm DOUBLE PRECISION")

    #Calculate area of study region from vector file
    grass.run_command("v.to.db",
                      map= study_area,
                      option = "area",
                      col = "area_sqkm",
                      unit = "k")

    # Get area of study region. This will need to be changed to the specific map name at a later
date.
    Study_area_km = grass.read_command("v.db.select",
                                       flags = "c",
                                       map = study_area,
                                       layer = 1,
                                       columns = "area_sqkm")

    Study_area_km = Study_area_km[:-2]
    Study_area_km = float(Study_area_km)

    # Calculated the required number of landslides
    num_slides = (slide_density * Study_area_km)
    num_slides = int(num_slides)
    print "The number of landslides dropped should be " + str(num_slides)

    # Commence a loop. Each iteration of the loop generates one synthetic susceptibility map
    and models its road network impact.
    iteration = 0
    while iteration < num_iterations:
        iter_name_start = "iter" + str(iteration) + "_start_time"
        iter_name_start = time.time()

        print "***************************************"
        print "ITERATION NUMBER:"
        print iteration

        #Optional: Randomly buffer sections of the road network
        grass.run_command("v.extract",
            input= "motorway_roads_split_cats",
            output = "motorway_roads_rand_selection",
            random = motorway_num_sections_to_be_buffered)

        grass.run_command("v.buffer",
            input= "motorway_roads_rand_selection",
            output = "motorway_roads_rand_selection_buffered",
            layer = 1,
            distance = motorway_bufwidth,
            minordistance = motorway_bufwidth)

        grass.run_command("v.extract",
            input= "secondary_roads_split_cats",
            output = "secondary_roads_rand_selection",
            random = secondary_num_sections_to_be_buffered)

        grass.run_command("v.buffer",
            input= "secondary_roads_rand_selection",
            output = "secondary_roads_rand_selection_buffered",
            layer = 1,
            distance = secondary_bufwidth,
            minordistance = secondary_bufwidth)

        grass.run_command("v.extract",
            input= "local_roads_split_cats",
            output = "local_roads_rand_selection",
            random = local_num_sections_to_be_buffered)
        grass.run_command("v.buffer",
            input= "local_roads_rand_selection",
```

```
        output = "local_roads_rand_selection_buffered",
        layer = 1,
        distance = local_bufwidth,
        minordistance = local_bufwidth)

    grass.run_command("v.extract",
        input= "minor_roads_split_cats",
        output = "minor_roads_rand_selection",
        random = minor_num_sections_to_be_buffered)

    grass.run_command("v.buffer",
        input= "minor_roads_rand_selection",
        output = "minor_roads_rand_selection_buffered",
        layer = 1,
        distance = minor_bufwidth,
        minordistance = minor_bufwidth)

    grass.run_command("v.patch",
        input =
        "motorway_roads_rand_selection_buffered,secondary_roads_rand_selection_buffer
        ed,local_roads_rand_selection_buffered,minor_roads_rand_selection_buffered",
        output = "patched_buffered_roads_all")

    grass.run_command("v.db.addtable",
        map = "patched_buffered_roads_all",
        columns = "susc double precision, roadsus int")

    # Adjust the value of susceptibility within those buffer zones
    grass.run_command("v.db.update",
        map = "patched_buffered_roads_all",
        column = "susc",
        value = 0)

    grass.run_command("v.overlay",
        ainput = study_area,
        binput = "patched_buffered_roads_all",
        operator = "or",
        output = "buffered_roads_all_studyreg")

    grass.run_command("v.db.addcolumn",
        map = "buffered_roads_all_studyreg",
        columns = "susc_multiplier integer")

    grass.run_command("v.db.update",
        map = "buffered_roads_all_studyreg",
        column = "susc_multiplier",
        value =1)

    grass.run_command("v.db.update",
        map = "buffered_roads_all_studyreg",
        column = "susc_multiplier",
        qcolumn = "b_susc",
        where = "b_susc = 0")

    grass.run_command("v.to.rast",
        input = "buffered_roads_all_studyreg",
        output = "buffered_roads_susc_multiplier",
        use = "attr",
        attrcolumn = "susc_multiplier")

    mapcalc_formula = "susc_map_roads_0 = " + str(susceptibility_map_rast) + "*"  +
    "buffered_roads_susc_multiplier"

    grass.mapcalc(mapcalc_formula)

    landslide_database = []
    tmpfile = grass.tempfile()
    # Create a temporary output file to store synthetic landslide attributes
    outf = file(tmpfile, 'w')
    print>> outf, 'N', 'E', 'L', 'W', 'randomseed', 'aream'
    for x in range(0, (num_slides*20)):
```

```
#Generate random Northing and Easting
Northing_center = random.uniform(S,N)
Easting_center = random.uniform(W,E)

#Generate random area from inverse gamma distribution
a= 1.4
location = -0.000132
scale = 0.00128
areakm2 = invgamma.rvs(a, location, scale)
if areakm2<0:
     continue
#Convert area to square meters from square kilometers
aream = (areakm2*1000000)

#Generate length to width ratio based on area
if aream<100:
                    a2 = 1.64
                    sc = 1.29
                    loc = 1.24
if aream >=100 and aream < 200:
                    a2 = 1.95
                    sc = 1.06
                    loc = 1.23
if aream >=200 and aream < 400:
                    a2 = 4.14
                    sc= 4.04
                    loc = 0.92
if aream >=400 and aream < 800:
                    a2 = 4.15
                    sc = 5.36
                    loc = 0.8
if aream >=800 and aream < 1600:
                    a2 = 3.52
                    sc = 5.75
                    loc = 0.68
if aream >=1600 and aream < 3200:
                    a2 = 2.94
                    sc = 6.64
                    loc = 0.5
if aream >=3200 and aream < 6400:
                    a2 = 2.53
                    sc = 8.64
                    loc = 0.19
if aream >=6400 and aream < 12800:
                    a2 = 2.95
                    sc = 9.37
                    loc = 0.32
if aream >=12800 and aream < 25600:
                    a2 = 3.39
                    sc = 14.81
                    loc =-0.18
if aream >=25600:
                    a2 = 1.86
                    sc = 5.87
                    loc = 0.72
lwratio = invgamma.rvs(a2, loc, sc)

#If we generate a lw ratio less than 1, replace with an 'average' value.
if lwratio<1:
                    lwratio = 1.8

scalefac = (aream/pi)
#Calculate the length of the long side of the landslide (f)
radiusL = math.sqrt(scalefac * lwratio)
#Calculate the length in cells of the short side of the landslide (g) based on
dividing the area by the longside (f)
radiusW = (radiusL/lwratio)

#Generate a random uniform value between 0 and 1 which is used to decide whether
a landslide is generated or not based on susceptibility
```

```
                randomseed = random.uniform(0,1)

                # Store values in temporary output file
                print>> outf, Northing_center, Easting_center, radiusL, radiusW, randomseed,
aream
          outf.close()

          #create column headers
          columns = ['north double precision',
                'east double precision',
                'radiusL double precision',
                'radiusW double precision',
                'randomseed double precision',
                'aream double precision']

          # Create a vector map of randomly generated points. The accompanying table contains
          a landslide area and length to width ratio for each point.
          grass.run_command('v.in.ascii', cat = 0, x = 2, y = 1, separator = 'space',
                input = tmpfile,
                output = "random_points"+str(iteration),
                columns = columns,
                skip = 1,
                overwrite = True)

          # Select only those points that lie within the study area extent
          grass.run_command("v.select",
                               ainput="random_points"+str(iteration),
                               alayer=1,
                               atype= "point",
                               binput= study_area,
                               blayer=1,
                               btype= "area",
                               output= "points_in_area"+str(iteration),
                               operator="overlap")

          grass.run_command("v.db.addcolumn",
                              map = "points_in_area"+str(iteration),
                              columns = "prob_map double precision, asp_map double precision,
                              susc_map_cat integer, susc_map_asp integer")


          # Read the value from the susceptibility map (including computational extent) and write
          the value into the column "prob_map"
          grass.run_command("v.what.rast",
                               map = "points_in_area"+str(iteration),
                               raster = "susc_rast",
                               column = "prob_map")

          #Read the value from the aspect map and write into the column
          grass.run_command("v.what.rast",
                               map = "points_in_area"+str(iteration),
                               raster = aspect_map,
                               column = "asp_map")

          #Select only those points where susceptibility map >= randomseed
          grass.run_command("v.extract",
                               input = "points_in_area"+str(iteration),
                               type = "point",
                               where = "prob_map >= randomseed",
                               output = "set_points_drop"+str(iteration))

          # Associate information about the aspect at that location for each landslide
          grass.run_command("v.what.rast",
                               map = "set_points_drop"+str(iteration),
                               raster = "susc_rast_cat",
                               column = "susc_map_cat")

          grass.run_command("v.what.rast",
                               map = "set_points_drop"+str(iteration),
                               raster = "susc_rast_asp",
                               column = "susc_map_asp")
```

```
          set_points_drop_list = grass.read_command("v.category",
                                input = "set_points_drop"+str(iteration),
                                type = "point",
                                option = "print",
                                flags = "g")

          set_points_drop_list_a = set_points_drop_list.split('\n')
          num_points_drop = len(set_points_drop_list_a)

          #Verify that enough points have been dropped (i.e., we have not rejected too many
landslides)
          if num_points_drop < num_slides:
               print "On this iteration, we didn't have enough landslides: " + str(iteration)

               continue
          else:

               #Select the number of final points that we want to add a landslide area to. These
               are selected randomly
               grass.run_command("v.extract",
                                 input = "set_points_drop"+str(iteration),
                                 output = "final_points"+str(iteration),
                                 random = num_slides)

               # Read the category value for each final point
               category_vals = grass.read_command("v.category",
                                 input = "final_points"+str(iteration),
                                 type = "point",
                                 option = "print",
                                 flags = "g")

               category_vals_a = category_vals.split('\n')
               num_points = len(category_vals_a)
               patch_list = []

               for i in range(0, num_slides):
                    category_value = category_vals_a[i]

                    #Extract the individual point, which will become the landslide centroid
                    grass.run_command("v.extract",
                         input = "final_points"+str(iteration),
                         output = "point"+str(category_value),
                         cats = category_value,
                         overwrite = True)

                    #Read the major & minor axes and the rotation (aspect) of the landslide from
                    the point file
                    major = grass.read_command("v.db.select",
                                                    flags = "c",
                                                    map = "point"+str(category_value),
                                                    columns = "radiusL")

                    minor = grass.read_command("v.db.select",
                                                    flags = "c",
                                                    map = "point"+str(category_value),
                                                    columns = "radiusW")


                    aspect = grass.read_command("v.db.select",
                                                    flags = "c",
                                                    map = "point"+str(category_value),
                                                    columns = "asp_map")

                    #Add a buffer around that point to give the landslide an area
                    grass.run_command("v.buffer",
                                      overwrite = True,
                                      input = "point"+str(category_value),
                                      output = "ind_landslide" +str(category_value),
                                      distance = major,
                                      minordistance = minor,
```

```
                    angle = aspect,
                    flags = "t")

#Overlay the landslide with the susceptibility map to see if the landslide
crosses any slope units
grass.run_command("v.overlay",
                    ainput = "overlay_susc",
                    binput = "ind_landslide"+str(category_value),
                    operator = "and",
                    output = "ind_landslide_susc"+str(category_value))

#Read the category values for the polygons within the landslide (these
correspond to the category values of the slope unit polygons)
category_vals_indls_a = grass.read_command("v.db.select",
                        flags = "c",
                        map = "ind_landslide_susc"+str(category_value),
                        columns = "cat")

# Check if the landlside has been split into multiple polygons by overlaying
with slope units map
category_vals_indls_a = category_vals_indls_a.split('\n')
length_cat_vals_indls_a = len(category_vals_indls_a)

if length_cat_vals_indls_a>1:
     del category_vals_indls_a[length_cat_vals_indls_a - 1]

category_vals_indls = []
for p in category_vals_indls_a:
     a = int(p)
     category_vals_indls.append(a)

if len(category_vals_indls) ==1:

     grass.run_command("v.db.addcolumn",
          map = "ind_landslide"+str(category_value),
          columns = "final_area DOUBLE PRECISION")

     grass.run_command("v.to.db",
          map = "ind_landslide"+str(category_value),
          option = "area",
          units = "meters",
          columns = "final_area")

#If we only have one polygons within the landslide, we need do nothing, as
the landslide has not crossed a slope unit.
#If we have two or more polygons within the landslide, we need to assess
whether the landslide needs to be trimmed
if len(category_vals_indls) >=2:

     #Get the category value (corresponding to the susceptibility map) of
     the centroid from reading this from the point database
     grass.run_command("v.what.vect",
                                  map = "point"+str(category_value),
                                  column = "susc_map_cat",
                                  qmap =
                                  "ind_landslide_susc"+str(category_val
                                  ue),
                                  qcolumn = "cat")

     centroid_cat = grass.read_command("v.db.select",
                                       flags = "c",
                                       map =
          "point"+str(category_value),
               columns = "susc_map_cat")

     centroid_cat = centroid_cat.split('\n')
     del centroid_cat[1]
     centroid_cat =(centroid_cat[0])
     centroid_cat = int(centroid_cat)

     #Get the aspect of the centroid by reading from the database of the point
```

```
                        layer
                        centroid_asp = grass.read_command("v.db.select",
                                                          flags = "c",
                                                          map =
"point"+str(category_value),
                                                          columns = "susc_map_asp")

                        centroid_asp = str(centroid_asp)
                        centroid_asp = centroid_asp.split('\n')
                        del centroid_asp[1]
                        centroid_asp = int(centroid_asp[0])

                        #Convert aspect from 0 to 7 to radians, then polar coordinates
                        if centroid_asp == 0:
                                centroid_asp = 0
                                centx = 0
                                centy = 1
                        if   centroid_asp == 1:
                                centroid_asp = 0.785398
                                centx = 0.71
                                centy = 0.71
                        if centroid_asp == 2:
                                centroid_asp = 1.570796
                                centx = 1
                                centy = 0
                        if   centroid_asp == 3:
                                centroid_asp = 2.356194
                                centx = 0.71
                                centy = -0.71
                        if centroid_asp == 4:
                                centroid_asp = 3.141593
                                centx = 0
                                centy = -1
                        if   centroid_asp == 5:
                                centroid_asp = 3.926991
                                centx = -0.71
                                centy = -0.71
                        if centroid_asp == 6:
                                centroid_asp = 4.712389
                                centx = -1
                                centy = 0
                        if   centroid_asp == 7:
                                centroid_asp = 5.497787
                                centx=-0.71
                                centy =0.71

                        centx = float(centx)
                        centy = float(centy)

                    #If there are two polygons within the landslide, we only need to compare one
                    and the other
                    if len(category_vals_indls) ==2:

                        Get the category value of the 'other' polygon (not the polygon with the
                        centroid in it)
                        category_vals_indls.remove(centroid_cat)

                        other_poly_cat = str(category_vals_indls)
                        other_poly_cat = other_poly_cat.strip("[")
                        other_poly_cat= other_poly_cat.strip("]")

                        #Read the aspect of that polygon
                        other_poly_asp = grass.read_command("v.db.select",
                                                    flags = "c",
                                                    map =
"ind_landslide_susc"+str(category_value),
                                                    where = "cat =" + str(other_poly_cat),
                                                    columns = "a_aspect")

                        other_poly_asp = str(other_poly_asp)
                        other_poly_asp = other_poly_asp.split('\n')
```

```
                        del other_poly_asp[1]
                        other_poly_asp = int(other_poly_asp[0])

                        #Convert that aspect to radians and polar coordinates
                        if other_poly_asp == 0:
                             other_poly_asp = 0
                             x = 0
                             y = 1
                        if   other_poly_asp == 1:
                             other_poly_asp = 0.785398
                             x = 0.71
                             y = 0.71
                        if other_poly_asp == 2:
                             other_poly_asp = 1.570796
                             x = 1
                             y = 0
                        if   other_poly_asp == 3:
                             other_poly_asp = 2.356194
                             x = 0.71
                             y = -0.71
                        if other_poly_asp == 4:
                             other_poly_asp = 3.141593
                             x = 0
                             y = -1
                        if   other_poly_asp == 5:
                             other_poly_asp = 3.926991
                             x = -0.71
                             y = -0.71
                        if other_poly_asp == 6:
                             other_poly_asp = 4.712389
                             x = -1
                             y = 0
                        if   other_poly_asp == 7:
                             other_poly_asp = 5.497787
                             x=-0.71
                             y =0.71

                        x = float(x)
                        y = float(y)

                        #Calculate the difference in aspects between the two polygons
                        asp_dif = math.sqrt((math.pow((centx-x),
2))+(math.pow((centy-y),2)))

                        #If the difference is greater than the threshold, select only the
                        centroid polygon to output as the final landslide area
                        if asp_dif>1.42:
                                  grass.run_command("v.extract",
                                       input = "ind_landslide_susc"+str(category_value),
                                       output = "ind_landslide"+str(category_value),
                                       where = "cat =" + str(centroid_cat),
                                       overwrite = True)

                                  grass.run_command("v.db.droptable",
                                       map = "ind_landslide"+str(category_value),
                                       flags = "f")

                                  grass.run_command("db.copy",
                                       from_table = "public.point" +str(category_value),
                                       to_table = "ind_landslide"+str(category_value))

                                  grass.run_command("v.db.connect",
                                       map = "ind_landslide"+str(category_value),
                                       table = "ind_landslide" +str(category_value))

                                  grass.run_command("v.db.update",
                                       map = "ind_landslide" + str(category_value),
                                       column = "cat",
                                       value = str(centroid_cat))

                                  grass.run_command("v.db.addcolumn",
```

```
                                        map = "ind_landslide"+str(category_value),
                                        columns = "final_area DOUBLE PRECISION")

                                   grass.run_command("v.to.db",
                                        map = "ind_landslide"+str(category_value),
                                        option = "area",
                                        units = "meters",
                                        columns = "final_area")

                                   print "For point " + str(category_value) + ", only the
                                   centroid was selected"
                              else:
                                   grass.run_command("v.db.droptable",
                                        map = "ind_landslide"+str(category_value),
                                        flags = "f")

                                   grass.run_command("db.copy",
                                        from_table = "public.point" +str(category_value),
                                        to_table = "ind_landslide"+str(category_value))

                                   grass.run_command("v.db.connect",
                                        map = "ind_landslide"+str(category_value),
                                        table = "ind_landslide" +str(category_value))

                                   grass.run_command("v.db.addcolumn",
                                        map = "ind_landslide"+str(category_value),
                                        columns = "final_area DOUBLE PRECISION")

                                   grass.run_command("v.to.db",
                                        map = "ind_landslide"+str(category_value),
                                        option = "area",
                                        units = "meters",
                                        columns = "final_area")

                         #If there are more than two polygons in the individual landslide, we need
                         to assess all of them
                         if len(category_vals_indls) >2:

                              #category_vals_indls.remove(centroid_cat)
                              all_poly_cats = str(category_vals_indls)
                              all_poly_cats_a = all_poly_cats.replace(","," or cat =")
                              all_poly_cats_b = all_poly_cats_a.strip("[")
                              all_poly_cats_c = all_poly_cats_b.strip("]")

                              #Get the aspects for each polygon within the landslide
                              polygons_aspects_a = grass.read_command("v.db.select",
                                                       flags = "c",
                                                       map =
"ind_landslide_susc"+str(category_value),
                                                       where = "cat =" + str(all_poly_cats_c),
                                                       columns = "a_aspect")

                              polygons_aspects_a = polygons_aspects_a.split('\n')
                              del polygons_aspects_a[(len(polygons_aspects_a) - 1)]

                              polygons_aspects = []
                              for p in polygons_aspects_a:
                                   a = int(p)
                                   polygons_aspects.append(a)

                              #Create a numpy array where column 1 is the category values and column
                              two is the aspects for each polygon
                              polygon_library = numpy.array([category_vals_indls,
polygons_aspects])

                              #compatible_polygon_cats = []
                              x_col = []
                              y_col = []

                              #For each polygon, convert its aspect to radians and polar coordinates
                              and assess the difference with the centroid aspect.
```

```
                    for j in range(0, len(category_vals_indls)):

                         other_poly_asp = polygon_library[1][j]
                         if other_poly_asp == 0:
                              other_poly_asp = 0
                              x = 0
                              y = 1
                         if   other_poly_asp == 1:
                              other_poly_asp = 0.785398
                              x = 0.71
                              y = 0.71
                         if other_poly_asp == 2:
                              other_poly_asp = 1.570796
                              x = 1
                              y = 0
                         if   other_poly_asp == 3:
                              other_poly_asp = 2.356194
                              x = 0.71
                              y = -0.71
                         if other_poly_asp == 4:
                              other_poly_asp = 3.141593
                              x = 0
                              y = -1
                         if   other_poly_asp == 5:
                              other_poly_asp = 3.926991
                              x = -0.71
                              y = -0.71
                         if other_poly_asp == 6:
                              other_poly_asp = 4.712389
                              x = -1
                              y = 0
                         if   other_poly_asp == 7:
                              other_poly_asp = 5.497787
                              x=-0.71
                              y =0.71
                         x_col.append(x)
                         y_col.append(y)

                    polygon_library = numpy.vstack([polygon_library, x_col])
                    polygon_library = numpy.vstack([polygon_library, y_col])

                    for j in range(0,len(polygon_library[0])):
                              xinit = polygon_library[2][j]
                              yinit = polygon_library[3][j]
                              difference = []
                              for p in range(0,len(polygon_library[0])):
                                   xcomp = polygon_library[2][p]
                                   ycomp = polygon_library[3][p]
                                   asp_dif = math.sqrt((math.pow((xinit-xcomp),
                                   2))+(math.pow((yinit-ycomp),2)))
                                   difference.append(asp_dif)

                              polygon_library = numpy.vstack([polygon_library,
difference])

                    #Create a table that lists the boundaries of all the polygons and which
                    polygon categories are to the left or right of that boundary line
                    grass.run_command("v.category",
                         input = "ind_landslide_susc"+str(category_value),
                         option="add",
                         out="polygons_neighbors"+str(category_value),
                         layer=2,
                         type="boundary",
                         overwrite = True)

                    grass.run_command("v.db.addtable",
                         map = "polygons_neighbors"+str(category_value),
                         layer = 2,
                         columns = "l_eft integer, r_ight integer")

                    grass.run_command("v.to.db",
```

```
        map = "polygons_neighbors"+str(category_value),
        option="sides",
        col="l_eft,r_ight",
        layer=2,
        type="boundary")

    #Report this so that we get a table in python. Column 1 is the category
    value of that boundary line, column two is the category value of poylgons
    to the left of that line and column three is the categeory value of
    polygons to the right of that line
    neighbor_list = grass.read_command("v.report",
        map = "polygons_neighbors"+str(category_value),
        layer =2,
        option = "length",
        units = "meters")

    all_compatible_polygons = [centroid_cat]
    neighbor_list_cat_a = neighbor_list.split('\n')
    num_neighbors =(len(neighbor_list_cat_a))-1

    neighbor_set = []
    for x in range(1, num_neighbors):
        d = neighbor_list_cat_a[x].split("|")
        e = map(float, d)
        e.remove(e[0])
        neighbor_set.append(e)
    neighbor_set = numpy.array(neighbor_set)
    neighbor_set = neighbor_set[neighbor_set[:,0] != -1]
    neighbor_set = neighbor_set[neighbor_set[:,1] != -1]

    all_difs = []
    for j in neighbor_set:
        left = j[0]
        right = j[1]
        col = left -1
        row = right + 3
        dif = polygon_library[row][col]
        j[2] = dif
        all_difs.append(dif)

    if max(all_difs)<1.43:

        grass.run_command("v.db.droptable",
                  map = "ind_landslide"+str(category_value),
                  flags = "f")
        grass.run_command("db.copy",
                  from_table = "public.point" +str(category_value),
                  to_table = "ind_landslide"+str(category_value))

        grass.run_command("v.db.connect",
                  map = "ind_landslide"+str(category_value),
                  table = "ind_landslide" +str(category_value))

        grass.run_command("v.db.addcolumn",
                  map = "ind_landslide"+str(category_value),
                  columns = "final_area DOUBLE PRECISION")

        grass.run_command("v.to.db",
                  map = "ind_landslide"+str(category_value),
                  option = "area",
                  units = "meters",
                  columns = "final_area")

    if max(all_difs)>1.43:

        #Find the polygon category values that neighbor the centroid
        import numpy as np
        neighboring_parcels_centroid_a = neighbor_set[neighbor_set[:,0]
        == centroid_cat]
        neighboring_parcels_centroid_b = neighbor_set[neighbor_set[:,1]
        == centroid_cat]
```

```
                                all_centroid_neighbors =
                                numpy.vstack([neighboring_parcels_centroid_a,neighboring_parce
                                ls_centroid_b])

                                final_centroid_neighbors = []
                                for g in all_centroid_neighbors:
                                     if g[2]<1.43:
                                          final_centroid_neighbors.append(g[0:2])

                                final_centroid_neighbors =
numpy.array(final_centroid_neighbors)
                                final_centroid_neighbors = final_centroid_neighbors.flatten()
                                final_centroid_neighbors = set(final_centroid_neighbors)
                                j = [centroid_cat]
                                final_centroid_neighbors = filter(lambda a: a!= j,
                                final_centroid_neighbors)

                                all_compatible_polygons = all_compatible_polygons +
                                final_centroid_neighbors

                                for s in final_centroid_neighbors:
                                     final_centroid_neighbors = set(final_centroid_neighbors)
                                     j = [centroid_cat]
                                     final_centroid_neighbors = filter(lambda a: a!= j,
                                     final_centroid_neighbors)
                                     neighbors_of_s_a = neighbor_set[neighbor_set[:,0] == s]
                                     neighbors_of_s_b = neighbor_set[neighbor_set[:,1] == s]
                                     neighbors_of_s = numpy.vstack([neighbors_of_s_a,
                                     neighbors_of_s_b])

                                     for g in neighbors_of_s:
                                          if g[2]<1.43:
                                               final_centroid_neighbors.append(g[0])
                                               final_centroid_neighbors.append(g[1])

                                final_centroid_neighbors.append(centroid_cat)
                                polygons_to_merge = set(final_centroid_neighbors)
                                polygons_to_merge = list(polygons_to_merge)
                                polygons_to_merge = str(polygons_to_merge)
                                polygons_to_merge_a = polygons_to_merge.replace(",","," or cat =")
                                polygons_to_merge_b = polygons_to_merge_a.strip("[")
                                polygons_to_merge_c = polygons_to_merge_b.strip("]")

                                polygons_to_merge = polygons_to_merge_c

                                grass.run_command("v.extract",
                                     input = "ind_landslide_susc"+str(category_value),
                                     output = "ind_landslide_b"+str(category_value),
                                     where = "cat = " + str(polygons_to_merge),
                                     flags = "d")

                                grass.run_command("v.db.addcolumn",
                                     map = "ind_landslide_b"+str(category_value),
                                     columns = "dissolve_cat int")

                                grass.run_command("v.db.update",
                                     map ="ind_landslide_b"+str(category_value),
                                     column = "dissolve_cat",
                                     value = 1)

                                grass.run_command("v.dissolve",
                                     input = "ind_landslide_b"+str(category_value),
                                     output = "ind_landslide"+str(category_value),
                                     column = "dissolve_cat",
                                     layer = 1,
                                     overwrite= True)

                                grass.run_command("v.db.droptable",
                                          map = "ind_landslide"+str(category_value),
                                          flags = "f")
```

```
                              grass.run_command("db.copy",
                                   from_table = "public.point" +str(category_value),
                                   to_table = "ind_landslide"+str(category_value))

                              grass.run_command("v.db.connect",
                                   map = "ind_landslide"+str(category_value),
                                   table = "ind_landslide" +str(category_value))

                              grass.run_command("v.db.update",
                                             map = "ind_landslide" + str(category_value),
                                             column = "cat",
                                             value = 1)

                              grass.run_command("v.db.addcolumn",
                                        map = "ind_landslide"+str(category_value),
                                        columns = "final_area DOUBLE PRECISION")

                              grass.run_command("v.to.db",
                                        map = "ind_landslide"+str(category_value),
                                        option = "area",
                                        units = "meters",
                                        columns = "final_area")

                         patch_list.append('ind_landslide'+str(category_value)+",")

                    # Exit the loop looking at landslides one by one, and patch individual landslide
                    vector files into one inventory
                    len_patch_list = len(patch_list)

                    if num_slides > 100:
                         num_patches = int(len_patch_list/100)
                         super_patch_list = []

                         for i in range(0, num_patches+1):
                              patch_list_a  = patch_list[(i*100):((i*100)+100)]
                              patch_list_2 = ''.join(patch_list_a)

                              if len(patch_list_a) > 1:
                                   grass.run_command("v.patch",
                                                  input = ( patch_list_2),
                                                  output =
"landslide_inventory_patch"+str(i),
                                                  flags = "e",
                                                  overwrite = True)


          super_patch_list.append("landslide_inventory_patch"+str(i)+",")
                              else:
                                   super_patch_list.append(patch_list_2)

                         super_patch_list_2 = ''.join(super_patch_list)
                         print super_patch_list
                         grass.run_command("v.patch",
                                        input = ( super_patch_list_2),
                                        output = "landslide_inventory_uncleaned"+str(iteration),
                                        flags = "e")
                    else:
                         patch_list_2 = ''.join(patch_list)

                         grass.run_command("v.patch",
                                                  input = ( patch_list_2),
                                                  output =
                                                  "landslide_inventory_uncleaned"+str(iteration),
                                                  flags = "e",
                                                  overwrite = True)

                    #Clean the topology of the final vector map
                    grass.run_command("v.clean",
                                             input =
"landslide_inventory_uncleaned"+str(iteration),
```

```
                            output = "landslide_inventory"+str(iteration),
                            tool = "bpol",
                            overwrite = "true")

    grass.run_command("v.build",
                            map = "landslide_inventory"+str(iteration))

    patch_list = []

    #Remove files from this iteration
    grass.run_command("g.mremove",
                            vect = "point*",
                            flags = "f")
    grass.run_command("g.mremove",
                            vect = "ind_landslide*",
                            flags = "f")
    grass.run_command("g.mremove",
                            vect = "final_points*",
                            flags = "f")
    grass.run_command("g.mremove",
                            vect = "random_points*",
                            flags = "f")
    grass.run_command("g.mremove",
                            vect = "set_points_drop*",
                            flags = "f")
    grass.run_command("g.mremove",
                            vect = "ind_landslide_susc*",
                            flags = "f")
    grass.run_command("g.mremove",
                            vect = "polygons_neighbors*",
                            flags = "f")
    grass.run_command("g.mremove",
                            vect = "landslide_inventory_uncleaned*",
                            flags = "f")
    grass.run_command("g.mremove",
                            vect = "set_points_drop*",
                            flags = "f")
    grass.run_command("g.mremove",
                            vect = "landslide_inventory_patch*",
                            flags = "f")

    #Overlay inventory with road network

    grass.run_command("v.overlay",
                             overwrite = True,
                             ainput = roads,
                             atype = "line",
                             binput = "landslide_inventory"+str(iteration),

                             operator = "and",
                             output = "road_blocks"+str(iteration))
    print "Overlaid landslide inventory with road map"

    grass.run_command("v.db.addcolumn",
                             map="road_blocks"+str(iteration),
                             columns="length double precision")
        grass.run_command("v.db.addcolumn",
                             map="road_blocks"+str(iteration),
                             columns="run int")
    grass.run_command("v.db.update",
                             map="road_blocks"+str(iteration),
                             column="run",
                             value = iteration)
    grass.run_command("v.to.db",
                             map = "road_blocks"+str(iteration),
                             option = "length",
                             units = "meters",
                             columns = "length")

    #Overlay inventory with road network plus buffer of 50 m
    grass.run_command("v.overlay",
```

```
                                        overwrite = True,
                                        ainput = buffered_roads50,
                                        atype = "area",
                                        binput = "landslide_inventory"+str(iteration),

                                        operator = "and",
                                        output = "road_blocks_within50"+str(iteration))

                grass.run_command("v.db.addcolumn",
                                        map="road_blocks_within50"+str(iteration),
                                        columns="length double precision")
                grass.run_command("v.db.addcolumn",
                                        map="road_blocks_within50"+str(iteration),
                                        columns="run int")
                grass.run_command("v.db.update",
                                        map="road_blocks_within50"+str(iteration),
                                        column="run",
                                        value = iteration)
                grass.run_command("v.to.db",
                                        map = "road_blocks_within50"+str(iteration),
                                        option = "length",
                                        units = "meters",
                                        columns = "length")

                #Overlay inventory with road network plus buffer of 100 m
                grass.run_command("v.overlay",
                                        overwrite = True,
                                        ainput = buffered_roads100,
                                        atype = "area",
                                        binput = "landslide_inventory"+str(iteration),

                                        operator = "and",
                                        output = "road_blocks_within100"+str(iteration))

                grass.run_command("v.db.addcolumn",
                                        map="road_blocks_within100"+str(iteration),
                                        columns="length double precision")
                     grass.run_command("v.db.addcolumn",
                                        map="road_blocks_within100"+str(iteration),
                                        columns="run int")
                grass.run_command("v.db.update",
                                        map="road_blocks_within100"+str(iteration),
                                        column="run",
                                        value = iteration)
                grass.run_command("v.to.db",
                                        map = "road_blocks_within100"+str(iteration),
                                        option = "length",
                                        units = "meters",
                                        columns = "length")

                #Add one to the iteration counter and start the next loop
                iteration = iteration + 1

if __name__ == "__main__":
     options, flags = grass.parser()
     atexit.register(cleanup)
     sys.exit(main())
```